

R E P O R T R E S U M E S

ED 002 967

24

A BASIC RESEARCH PROGRAM ON READING.

BY- LEVIN, HARRY AND OTHERS

CORNELL UNIV., ITHACA, N.Y.

REPORT NUMBER CRP-639

PUB DATE

63

EDRS PRICE MF-\$0.54 HC-\$15.72 393P.

DESCRIPTORS- *READING RESEARCH, *READING DEVELOPMENT, *READING PROCESSES, *LEARNING EXPERIENCE, *EDUCATIONAL EXPERIMENTS, WRITING, SPELLING, PHONETICS, EXPERIMENTAL PROGRAMS, ITHACA, NEW YORK

A THEORETICAL ANALYSIS OF THE READING PROCESS, BOTH PSYCHOLOGICAL AND LINGUISTIC, AND THE STUDIES DEVELOPING FROM THE ANALYSIS FORMED THE BASIS FOR THIS RESEARCH PROGRAM ON READING. A RESEARCH GROUP WAS ORGANIZED AS A LOOSE FEDERATION WITH EACH OF THE PRINCIPAL INVESTIGATORS PURSUING HIS OWN IDEAS AND TAKING RESPONSIBILITY FOR THEM. THE COMPLETED STUDIES (22 IN NUMBER) ARE PRESENTED IN THE REPORT, FALLING INTO SEVERAL GROUPS. GROUP ONE COVERED THE ACQUISITION OF READING SKILL THROUGH SUCH ACTIVITIES AS (1) LEARNING TO SPEAK ONE'S LANGUAGE, (2) DISCRIMINATION OF LETTERS, (3) LEARNING TO DECODE GRAPHIC SYMBOLS TO THEIR APPROPRIATE SPEECH SYMBOLS, AND (4) FORMING OR DISCOVERING HIGHER UNITS FOR RAPID READING. GROUP TWO DISCUSSED (1) THE RELATIONSHIP BETWEEN READING AND WRITING, (2) THE LEARNING OF GRAPHEME-TO-PHONEME CORRESPONDENCES, AND (3) THE SOURCES OF CONFUSION ERRORS IN THE RECOGNITION OF WORD FORMS. GROUP THREE DISCUSSED THE ROLE PLAYED BY VARIOUS TYPES OF STIMULUS ATTRIBUTES WHEN THEY ARE PRESENT AS COMPETING CUES IN TASKS REQUIRING A CHILD TO MAKE COMPARATIVE SIMILARITY JUDGMENTS. GROUP FOUR ANALYZED SOME OF THE IMPLICATIONS OF LINGUISTICS FOR THE PROCESS OF READING. GROUP FIVE COVERED RESEARCH AIMED AT THE EMPIRICAL DEFINITIONS OF UNITS IN LANGUAGE. (AW)

ED 002967

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

A BASIC RESEARCH PROGRAM ON READING

Cooperative Research Project No. 639

Harry Levin
Eleanor J. Gibson
Alfred L. Baldwin
James J. Gibson
Charles F. Hockett
Henry N. Ricciuti
George J. Suci

Cornell University

Ithaca, New York

1963

The research reported herein was supported by the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare

CONTENTS

Preface

A Developmental Study of the Discrimination of Letter-Like Forms
Eleanor J. Gibson, James J. Gibson, Anne D. Pick, and Harry Osser

An Analysis of Critical Features of Letters, Tested by a Confusion Matrix
Eleanor Gibson, Harry Osser, William Schiff, and Jesse Smith

Improvement in Visual Discrimination of Letter-Like Forms
Anne D. Pick

A Possible Effect of Learning to Write on Learning to Read
James J. Gibson and Harry Osser

Transfer of Word and Letter Training in Reading
Carol Bishop

The Role of Grapheme-Phoneme Correspondence in the Perception of Words
Eleanor J. Gibson, Anne Pick, Harry Osser, and Marcia Hammond

A Study of the Development of Grapheme-Phoneme Correspondences
Eleanor J. Gibson, Harry Osser, and Anne D. Pick

A comparison of Meaningfulness and Pronunciability as Grouping Principles in the Perception and Retention of Verbal Material
Eleanor J. Gibson, Carol H. Bishop, William Schiff, and Jesse Smith

The Effects of Motor Pretraining on Children's Association and Recognition Learning
Harry Levin

Writing as Pretraining for Association Learning
Harry Levin, John S. Watson, and Margaret Feldman

The Effect of Interpolated Writing on Paired-Associate Learning
Linda Johnson

The Learning of Variable Grapheme-to-Phoneme Correspondences
Harry Levin and John Watson

The Learning of Variable Grapheme-to-Phoneme Correspondences: Variations in the Initial Consonant Position
Harry Levin and John Watson

The Learning of Variable Grapheme-to-Phoneme Correspondences: Comparison of English and Spanish Speakers
Harry Levin, Esther Baum, and Susan Bostwick

The Use of Cues in Word Recognition
Gabrielle Edelman

CONTENTS (Continued)

Geometric Form and Detail as Determinants of Comparative Similarity
Judgments in Young Children
Henry N. Ricciuti

Selective Cue Utilization in Young Children's Comparative Similarity
Judgments of Pictorial Stimuli
Henry N. Ricciuti

The Elements of Graphonomy
Charles F. Hockett

Analysis of English Spelling
Charles F. Hockett

A Computer Program for Deriving Spelling-to-Sound Correlations
Richard L. Venezky

The Interruptability of Words in the Speech of Nursery School Children
A. L. Baldwin and Esther Baum

The Definition of Units in Language
George J. Suci

A BASIC RESEARCH PROGRAM ON READING

Preface

Four years ago, when work on this program was begun, we held a strong bias as regards the kind of work which we were equipped to do and which, we thought, was needed. Our bias was the conviction that our research should be basic; that it should stem from a theoretical analysis of the reading process, both psychological and linguistic, and that the studies developing from the analysis should be carried out with the most modern methods of behavioral science. The time was ripe, we thought, to contribute to the understanding of reading from the point of view of behavioral science, and to contribute to behavioral science by a study of the reading process. The search was to be for theories and correct principles, and the methods to be experimental techniques and careful description, developing in the laboratory "miniature" situations analogous to real ones.

The number of publications bemoaning the present status of reading instruction which have appeared during these past four years strengthened our faith that a new approach, based on theory and culminating in experimental tests of our hypotheses, was a good choice. We have been heartened, also, by the interest of psychologists and educators alike as soon as we had material ready to report at meetings and in preliminary written form. The research must stand on its own, but we feel that some new directions have been indicated. If we have helped lure the educator into the laboratory and the scientist into an educational problem, something at least will have been achieved.

The organization of the research group, an interdepartmental one, was a loose federation with each of the principal investigators pursuing his own ideas and taking sole responsibility for them. But it is safe to say that our group meetings and frequent discussions resulted in a mutually profitable exchange of ideas, as well as useful criticism of hypotheses and techniques.

The completed studies are presented in the report as individual papers, arranged in groups ordered according to the principal investigators and the work which each one took responsibility for. Some of the sub-projects were presented as dissertations by research assistants working on the project, and in these cases a large share of the responsibility was assumed by the candidate. Much that was creative and new came from these dissertations. The original group of investigators acknowledges a debt of gratitude not only for assistance from these people but for ideas as well.

A DEVELOPMENTAL STUDY OF THE DISCRIMINATION OF
LETTER-LIKE FORMS¹

Eleanor J. Gibson, James J. Gibson,
Anne D. Pick, and Harry Osser

Cornell University

It has generally been assumed in psychological analyses of the reading process that ability to discriminate letters is a prerequisite for reading and that children of 6 years, when they begin to read, can in fact distinguish one letter from another with reasonable accuracy. The relevant literature is summarized by Vernon (1957), pp. 20 ff.). Some studies have used drawing as a criterion, some naming, and some matching. No one study has traced the development of letter differentiation as related to those dimensions or features of letters which are critical for the task and which may present more or less difficulty.

A method of studying this problem was suggested by an experiment of Gibson and Gibson (1955) in which children of two age groups and adults identified "scribbles" which were systematically varied on three separate dimensions (number of coils, compression, and orientation). The scribbles were somewhat comparable to strokes used in cursive writing. The results showed that the difficulty of the task was greater the younger the S, that errors varied in number with the dimension varied, and that systematic variation along at least three dimensions was possible. It also showed that confusion errors (primary generalization) provided an effective criterion for measuring ability to differentiate line drawings of this type.

1. Published in Jour. Comp. Physiol. Psychol., 1962, 55, 897-905.

The authors wish to express their gratitude to the principal and teachers of South Hill School, Ithaca, for their interest and generous cooperation.

The present experiment was designed to study the development of the ability to discriminate visually a set of letter-like forms in children 4 through 8 years of age. The aim was not merely quantitative comparison of different age levels, but primarily a qualitative developmental study of types of error as related to certain critical features of letters.

To secure information on qualitative changes, the plan was adopted of constructing specified transformations for each of a group of standard letter-like forms. The transformations were chosen on an intuitive basis with regard to the distinctive features of letters as a set. Three types of transformation were included which were considered critical for discriminating printed letters, and one which was not.

Letter-like forms were selected, rather than actual letters, in order to keep specific experience with the forms as equal as possible. Matching, rather than identification, was the task, since simple discrimination (seeing a difference) was the criterion desired.

METHOD

Materials

Construction of standard forms. An analysis was made of actual letters (printed capitals, upper case, of the simple type customarily used in primary texts) in terms of number of strokes, straight vs. curved lines, angles, open vs. closed forms, symmetry, etc. This procedure provided a set of "rules" which describe generally the construction of letters. New forms were generated which follow the same constraints. A large number of forms resulted from which 12 were chosen as standards for the experiment. Of these half were symmetrical and half asymmetrical, half open and half closed, some combined straight and curved lines, some were composed only of straight lines, some only of curves, and the number of strokes varied from two to four.

Construction of transformations. The transformations chosen were as follows: three degrees of transformation of line to curve or curve to line (1, 2, or 3); five transformations of rotation or reversal, i.e., 45^0 rotation, 90^0 rotation, 180^0 rotation, right-left reversal, and up-down reversal; two perspective transformations, a 45^0 slant left and a 45^0 backward tilt; and two topological transformations, a break and a close.

The transformations were constructed by a draftsman, tracing from the standard whenever possible so that no change other than the intended one would be introduced. The perspective transformations were made by photographing each standard at the desired slant or tilt. The copies to be used for the experiment were reproduced by a photo-static process from the master copies. They were then mounted on 1 1/4-inch square cards and covered with plastic.²

It will be noted that half the forms (the symmetrical ones) lacked two transformations, since either the right-left reversal or up-down reversal was identical with the standard and the 180^0 rotation was identical with the other. The cells in these cases were filled with the standard.

Procedure

The task. The discrimination task required S to match a standard with an identical form, in the following manner: A standard form was placed in the center of the top row of an apparatus which might be thought of as a matrix board. It was constructed of black-painted wood and contained four slotted rows below the top one, tilted so that S's angle of regard would be the same for each row. Each row

2. Illustrations of all the forms may be seen in the published article.

held 13 cards. In any single row, a given standard and its own transformations were placed in a random order. Three boards were available, so that 12 rows, one for each standard and its transformations could be set up before the experiment began. For the forms which lacked two transformations, three copies of the standard were mixed with the transformations. The others had one copy each of the standard in the row. This unequal number of exact matches per row had the advantage of preventing S from finding one and then stopping without scrutinizing the other cards.

After E had put the appropriate standard in the center of the top row, S went through a given row, searching for any form which was "exactly like the standard." When he found one, he removed it and handed it to E and continued until he had scanned the entire row. The display on the board at any given time had four filled rows, except for spaces left where S had removed a presumed match. All Ss matched for all 12 standards in counterbalanced orders.

The instructions were given very explicitly, so as to make absolutely clear that only an exact match was wanted. A demonstration was given first with very large sample forms (real letters) which included two reversals. The E asked if they were the same and corrected S if he responded incorrectly. Repetition and other terms such as "equal" or "exactly like" were used if necessary. Then S was given a practice row on the matrix board, again with real letters, but with a standard at top and filled board so that conditions simulated the final task. The S scanned across the row indicated by E, matching the standard, and E corrected his errors and explained the task again, if needed. If S lost his place or seemed not to be looking at every

letter, his regard was directed across the board by E's indicating each one in turn. His attention was recalled to the standard if errors were made. No correction was given after the main task was begun, but the younger children had their attention directed back to the row and were guided across if their gazes wandered. There was no time limit.

Scoring. When S withdrew a form from the board and gave it to E, it was filed in a box and later classified as a correct match or an error. An error could be of two kinds: (a) failure to recognize a standard or (b) a confusion error, that is, selection of a transformation as identical with a standard. In the latter case, errors were classified according to type of transformation.

Subjects

The Ss were 167 children aged four through eight. The Ss from kindergarten through third grade were obtained at the same school, where the experiment was run during school hours. When grades were grouped according to progress, samples were taken from each group. The younger children were obtained at three nursery schools, the Cornell University Nursery School, a church nursery school, and a public nursery school at a settlement house. Numbers varied somewhat for age groups, but each age group contained at least 24 Ss, in order to rotate twice through the order.

Validation Experiment

It seemed desirable to have a check on the validity of the "letter-likeness" of the forms as well as a replication of the transformations with other standards. For this reason, a second experiment was run

in which the standards were 12 real letters (Roman capitals, of the type used to formulate the rules in the first place). The 12 transformations were the same as those of the main experiment, and the material was constructed in the same way. The task and the apparatus duplicated the first experiment.

The Ss were the kindergarten children tested in the main experiment, except for two who were no longer available. There was a 2 month interval between experiments.

RESULTS

Order Effects

A check was made to see whether order (position of the standard in the series of 12) was a source of variance. By Tukey's test of differences among a group of means (see Ryan, 1960), no difference was significant. Order was therefore disregarded thereafter and the data were combined.

Pooled Errors by Age Groups

There were very few errors of omission (failure to select the true match)--these errors did not vary in any orderly way, as Table 1 shows; therefore, all further discussion will be concerned with confusion errors, that is, identifying a variable as a standard. Pooled mean errors of this latter type declined with age, as would be expected (see Table 1). The standard deviations indicate rather large individual differences, which also decreased with age except for the 6-year-olds, who were apparently an unusually heterogeneous group. The drop in mean error from 5 to 6 years, furthermore, was not as great as that between other age groups. The difference in mean error between boys and girls was small for all five groups, favoring the girls through 6 years and then the boys.

TABLE 1

Pooled Mean Errors by Age Group

Measure	Age				
	4	5	6	7	8
N	26	35	29	30	32
Omission errors	1.31	1.83	.41	.87	.94
Confusion errors					
Boys	61.69	40.42	37.76	24.53	16.63
Girls	54.54	38.31	36.25	27.85	22.88
M	58.12	39.46	37.14	25.97	19.75
SD _M	25.39	20.75	23.20	11.00	10.95

Analysis of Errors by Transformation

The mean errors for each type of transformation, presented in Table 2 for the five age groups, differ in frequency for different transformations as well as for age groups. There was a decrease in errors for all transformations as age increased, but some transformations were harder to discriminate from the standard than others, and rate of improvement varied with the type of transformation.

The errors for both the topological transformations--the changes of close and break--were few even for the 4-year olds and declined to almost zero for 8-year olds. For both perspective transformations, on the other hand, the errors started very high ($M = 9.88$ and 9.23 at 4 years, out of a possible 12), and were still high at 8 years.

The errors for rotations and reversals started high, but by 8 years had declined almost to zero.

The errors for changes of line to curve were relatively great at 4 years for the first degree transformation and dropped low by 8 years. For these transformations, mean error varied with the number of changes. Gradients of generalization for the three degrees of transformation can

be compared by age groups. Generalization was greater, the younger the child, except for the overlapping of 5- and 6-year olds, and in every case decreased as the degree of change increased.

TABLE 2
Mean Errors Made for Each Transformation by Age Groups

Transformation	Age Groups				
	4(N=25)	5(N=35)	6(N=29)	7(N=30)	8(N=32)
Curve to line (1)	5.85	4.06	4.00	2.53	1.28
Curve to line (2)	4.42	2.60	2.69	1.33	0.53
Curve to line (3)	3.04	1.46	1.76	0.60	0.31
450 rotation	5.19	2.14	1.79	0.53	0.78
900 rotation	4.31	1.48	1.28	0.03	0.34
Right-left reversal ^a	6.56	3.96	2.07	0.97	0.59
Up-down reversal ^a	6.47	3.55	2.44	1.56	1.08
1800 rotation ^a	5.24	2.74	1.10	0.14	0.38
Perspective, hor.	9.88	9.20	9.69	9.27	7.34
Perspective, ver.	9.23	8.97	9.31	8.20	6.81
Close	1.19	.69	0.83	0.43	0.31
Break	2.62	1.86	1.86	1.07	0.59

^a These figures have been corrected to allow for the fact that opportunities for error were less than for the other transformations.

The differences in development of discrimination depending on type of transformation are most easily seen in Figure 1. Errors were combined for each of the four transformation groups. The curves not only look different from one another but the differences are statistically significant at most points of comparison. Ryan's (1960) method for making multiple comparisons was applied to proportions of errors to test the differences between the four types of transformation. At age four, all the differences were significant, and at five, six, and seven, five of the six differences were significant. At eight, where errors were few, the differences were significant only between perspective transformations and the other three types.

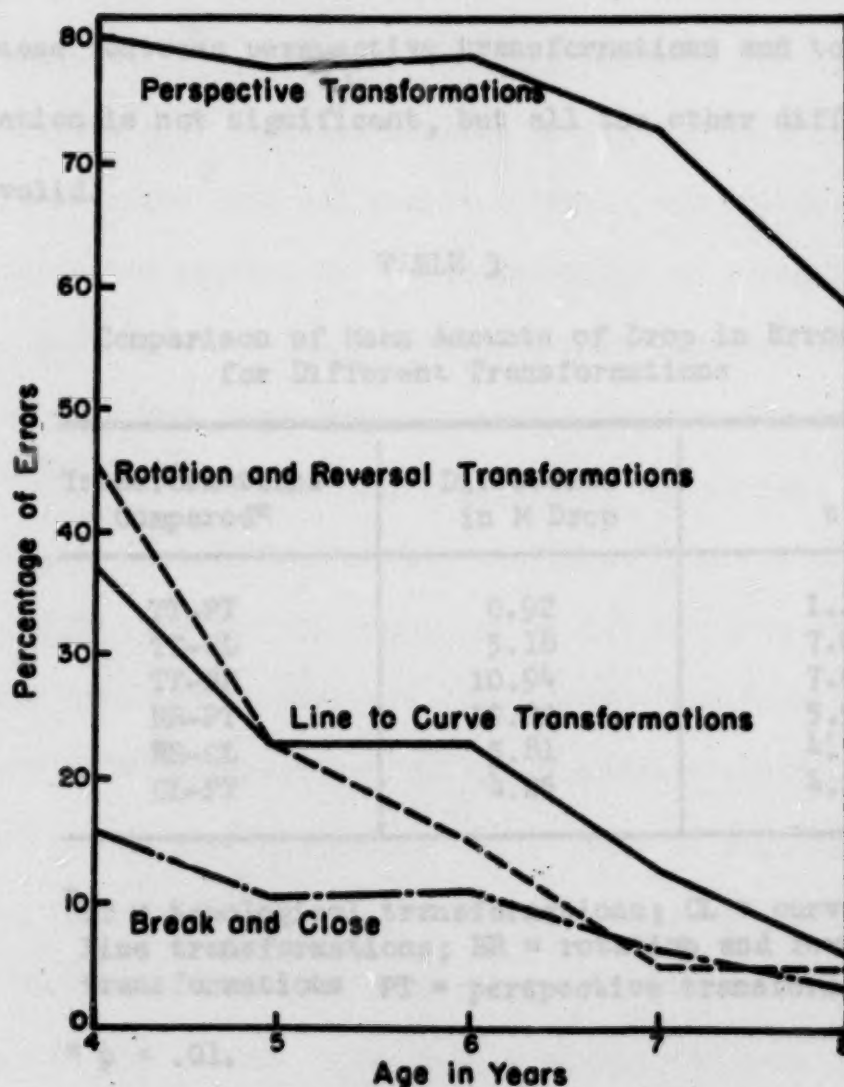


Figure 1. Developmental Error Curves for Four Types of Transformation.

The slopes of the four curves suggest that there is the greatest developmental change between 4 and 8 years in tendency to confuse rotation-reversals, with line-to-curve errors showing the next greatest drop, perspective errors next, and topological errors least. To test the differences in drop in error with age, the two youngest age groups were combined and the two oldest combined and the mean amount of drop computed for each type of transformation. Then "t" tests were made of differences between these mean drops for all six

transformation comparisons (see Table 3). The difference in error decrease between perspective transformations and topological transformation is not significant, but all the other differences in slope are valid.

TABLE 3
Comparison of Mean Amounts of Drop in Error
for Different Transformations

Transformations Compared ^a	Difference in M Drop	t
TT-PT	0.92	1.12
TT-CL	5.18	7.01*
TT-RR	10.94	7.03*
RR-PT	10.22	5.94*
RR-CL	5.81	4.57*
CL-PT	4.25	4.29*

^a TT = topological transformations; CL = curve to line transformations; RR = rotation and reversal transformations PT = perspective transformations.

* $p < .01$.

Correlations between Types of Error

With reference to the four curves of Figure 1, some justification should be given for pooling errors within transformation groups. There were several reasons for so doing: First, ad hoc or intuitive reasons, because the transformations combined were of the same geometrical types and because the curves, when drawn individually, resembled each other within transformation groups. Furthermore, the differences between curves of different transformation groups were overall significant. It was also possible to obtain a third kind of evidence, i.e., the intercorrelations between different kinds of transformation error.

Correlations were run between errors for each of the 12 transformations and every other, separately for each age level. This gave a 12 X 12 correlation matrix for each age level. Because space does not permit printing all the five matrices, the information yielded by the correlations is summarized in Table 4. The correlations within a transformation group, where there is more than one, have been averaged. So have the correlations between that transformation group and all others. For instance, the mean correlation between the three line-to-curve transformations is +.88 for the 4-year-olds, but the mean of the line-to-curve transformations with other types (rotations, reversals, perspective transformations and topological transformations) is lower, +.62. A similar comparison can be made at each age level for each transformation group. The N in Table 4 refers to the number of correlation coefficients which have been averaged for any mean in the column. (The S N is always the same for any age group.) Ranges are presented under the means in brackets to indicate the lowest and highest of the coefficients averaged in any mean.

It will be noted that in 18 of the 20 comparisons within and between transformation groups, the mean correlation is lower between different transformation types. In the two atypical cases (rotation-reversal and topological transformations at the 7-year level), the correlations are low and insignificant. The reason for this is that errors of these types had dropped to zero for nearly all Ss. The correlations for the 7- and 8-year olds are less meaningful than those at four, five, and six because of the

TABLE 4
Mean Correlations Within and Across Transformation Groups

Age	L to C		R & R		P.T.		T.T.	
	Within N=3	Across N=27	Within N=10	Across N=35	Within N=1	Across N=20	Within N=1	Across N=20
4	.88 (.85 to .91)	.62 (.32 to .81)	.77 (.63 to .91)	.54 (.28 to .81)	.73	.43 (.25 to .59)	.59	.52 (.25 to .75)
5	.76 (.71 to .83)	.55 (.29 to .85)	.74 (.61 to .87)	.43 (.01 to .73)	.68	.33 (.01 to .53)	.65	.52 (.28 to .85)
6	.90 (.87 to .93)	.54 (.37 to .88)	.78 (.64 to .95)	.42 (.18 to .66)	.85	.40 (.15 to .60)	.55	.48 (.15 to .88)
7	.70 (.66 to .77)	.22 (-.19 to .79)	.07 (-.20 to .38)	.14 (-.17 to .50)	.81	.26 (-.15 to .57)	-.19	.17 (-.19 to .79)
8	.58 (.45 to .73)	.46 (.04 to .73)	.77 (.60 to .97)	.34 (-.05 to .74)	.75	.35 (-.05 to .53)	.48	.18 (-.03 to .45)

lowness of the error scores for everything except perspective transformations. The topological transformations showed less contrast than the others, probably because they are not as "clean" transformations. The break or the close had to be inserted somewhat arbitrarily, so that configurational changes in addition to the desired one probably resulted in some instances.

The high intercorrelations between transformations of the same type furnish strong supportive evidence for the inner consistency of the types of features chosen for qualitative analysis of errors.

Comparison with Real Letters

A check on validity, both of the material used and the functioning of the different transformations, was furnished by the repetition of the experiment with 12 real letters and the same 12 transformations. Only the kindergarten group took part. In Table 5 the mean number of errors for each transformation can be compared for the letter-like forms and the letters. Errors were fewer for letters. This trend might be expected, since the children were familiar with letters (they could print their names) and they were 2 months older. But the correlations between transformations are significant in every case.

One can also ask whether the order of difficulty of transformations corresponds in the two cases. The rank-order correlation between the forms and the letters, ordering the transformations according to mean number of errors, is $+0.87$, showing that there was high correspondence.

Correlations within transformation types and across them were obtained for the letters, as they were for the forms. Mean r 's within transformation groups were $.83$, $.76$, $.73$, $.73$; between groups,

.56, .44, .38, .53. As was the case with the forms, the correlations within groups were considerably higher than across groups.

TABLE 5

Mean Errors by Transformation for
Letter-Like Forms and Letters,
and the Correlations
between Them

Transformation	M		Correlation between Forms and Letters
	Forms	Letters	
Curve to line (1)	4.33	4.25	.56
Curve to line (2)	2.78	1.98	.37
Curve to line (3)	1.60	1.15	.42
45° rotation	2.48	1.18	.84
90° rotation	1.90	1.18	.84
R-L reversal	3.33	2.25	.77
Up-Down reversal	2.93	1.03	.75
180° rotation	1.53	.65	.75
Perspective, hor.	9.38	7.85	.49
Perspective, vert.	8.95	7.13	.37
Close	.88	.68	.63
Break	1.85	1.23	.59

These results appear to confirm very satisfactorily the qualitative differences produced by different transformations in the first experiment. From this evidence, it is concluded that transformation type is a good predictor of confusion errors.

Comparison of Forms

Although characteristics of the individual standard forms were not a matter of primary interest for the experiment, it is possible to classify and compare them in certain respects (e.g., symmetrical vs. asymmetrical) to see if these qualities are associated with greater or less discriminability. The "perceptibility" of forms in relation to such contour characteristics as symmetry, continuity, etc., has

been studied by Fitts, Weinstein, Rapaport, Anderson, and Leonard (1956) and recently by Klemmer (1961). Undoubtedly, the role played (if any) by a dimension such as symmetry in "pattern difficulty" (to use Klemmer's term) would vary with the judgment required and with the set of other items also being judged. Under the conditions of the present experiment, symmetry did have a significant effect. Symmetrical standards were confused with their variants less often than asymmetrical ones were confused with theirs. The difference appeared in all five age groups ($p < .02$ in every case and $< .001$ in three cases), but the magnitude of the difference was very small. This finding agrees with the two studies mentioned.

Klemmer investigated, also, the effect of continuity and of a closed loop. All our standards were continuous, but since half the standards possessed a closed loop and half did not, they were compared for the latter characteristic. Here the results disagreed with Klemmer's; more errors were made when the standard possessed a closed loop. The difference was not significant at every age level, although it was in the same direction.

The straight-curved classification in our experiment was a three-way one, for four standards were all straight, four all curved, and four mixed. When these three types were compared, the all-curved standards were associated, at all age levels, with more errors than all straight. But the mixed figures occupied an ambiguous position with respect to them.

These "absolute" characteristics of a standard seem to us trivial as predictors of identifiability when compared with the transformations within the set to be judged. In other words, the influence

of similarity within a set (here the transformations in relation to the standard) seems to us more important than any characteristic of the standard as such.

DISCUSSION

The most obvious and hardly unexpected outcome of the experiment was that the visual discrimination of letter-like forms, using a matching procedure which requires a judgment of same or different, improves from age four to eight. What are the reasons for improvement? Better ability to keep the task in mind and follow it through cannot be wholly ruled out, but preliminary practice, instruction, and assistance in scanning when required rendered this factor minimal. Maturation of retinal processes is still continuing during this period, but acuity as such should have been adequate even for the 4-year-olds. It seems, rather, that children between four and eight learn something about letter-like forms which makes possible better discrimination even between ones they have not seen before.

Our results suggest that what they learn are the features or dimensions of difference which are critical for differentiating letters. Some are critical, but some are not, and the latter variations in letters must in fact be tolerated. Here it is useful to consider Roman Jakobson's concept of "distinctive features" of phonemes (Jakobson & Halle, 1956). Phonemes, like graphemes, must be differentiated from one another, but many variations are permissible without destroying identifiability; the same phoneme pattern is recognizable when it is delivered by a high-pitched or low-pitched voice, whispered, shouted, sung, and so on. Brown (1958) has referred to this as

"constancy of the phoneme." But certain features--the "distinctive" ones--are invariant and must be heard as such. It is our hypothesis that it is the distinctive features of grapheme patterns which are responded to in discrimination of letter-like forms. The improvement in such discrimination from four to eight is the result of learning to detect these invariants and of becoming more sensitive to them.

The 4-year-olds do, of course, discriminate the letter-like forms up to a point. They have had some experience with alphabet blocks and picture books. But, more important, some of their previous experience with solid objects could transfer to this new discrimination task. Solid objects also have invariant qualities and distinctive features, and a 4-year-old has learned what he needs to distinguish many of these, perhaps at the same time he began to distinguish the invariants which permit constancy of shape and size as he moves about in space. The question is, Why should the four types of transformation differ in difficulty at the outset and why do errors decrease at different rates?

The perspective transformations produced the greatest number of confusions in the 4-year-olds. Was this because the amount of physical change was small? We have no absolute measure for comparing degrees of change. Nevertheless a 45° slant is very appreciable; acuity is good for it with judgments of moving perspective transformations (Gibson & Gibson, 1957) in adult Ss. But there was no motion in the present experiment--i.e., no indication of a depth difference. Furthermore, perspective transformations are not distinctive features for object identification. On the contrary, these variations must be tolerated for shape constancy to occur. Therefore, no transfer from discriminating solid objects of the world can be expected for this

dimension, since the compression-like changes which result have never been critical for judging objects as same or different. Consequently, errors occurred at a very high rate when S was asked to match letter-like forms differing only by this type of transformation .

The decrease in these errors with age was also very slight because perspective transformations are no more distinctive features of letters than of objects. More or less compression must be tolerated in reading the same letter when the page is held at various angles to the line of sight.

At the other extreme are topological transformations--changes produced by breaking or closing a line. Differences of this type are not continuous, and they are critical for object discrimination. Piaget and Inhelder (1956) showed that children discriminate them with solid objects at a very early age. The difference is perceived as well in line drawings. The error curve for breaks and closes of lines began low initially and dropped, reaching zero at 8 years, because this feature is critical for distinguishing letters as well as objects (i.e., the difference between C and O).

Initial errors for reversals and rotations would also be expected to be high because transformations of this type are not critical for identification of objects as same or different. Rotation gives information about the position of an object. If habits formed for object-discrimination are used when the child begins to make graphic discriminations, confusions due to rotation are to be expected, and in fact occurred. A right-left reversal (mirror reversal) is equivalent to turning over a flat object from front to back and again is not significant, usually, for distinguishing solid objects. Although this curve

started high, it fell rapidly to near zero at 8 years, for the child learns during this period that transformations of rotation and reversal are significant for distinguishing graphic forms (i.e., the difference between M and W, C and U, d and b, p and b).

The error curves for line-to-curve transformations need a more complex interpretation. For solid forms, the equivalent transformation would be a deformation of a rigid object, and in that case a critical one. But for living things or plastic objects, elastic transformations of line to curve do occur (i.e., changes of facial expression) and are not significant for identifying an object as same or different; they indicate, rather, a change of state. Transfer to graphic discriminations is therefore equivocal. The amount of the change (one to three changes) was important here. With three changes the error curve was initially low, but with one it was high. The prediction for eventual discrimination, however, is clear. These changes are critical ones for graphemes (i.e., the difference between U and V), and the error curve, at eight, has dropped almost to zero.

The developmental error curves for the four types of transformation that were studied can, then, be interpreted successfully and convincingly by the hypothesis that the child has to learn the distinctive features of letters and that he does not start "cold" because of transfer from his already good ability to differentiate critical features of objects. It may occur to the reader that we have not dealt with all the features by which one letter differs from another. This is of course the case; the set of distinctive features of letters is greater than the set of variables chosen here. Also, the set of non-critical variations of a letter was not fully explored (size, for

instance). An attempt is at present being made to work out a fuller set of distinctive features for graphemes. When the set is available, predictions can be made and tested with a confusion matrix.

Another question of interest is how the distinctive features are detected by the child. Teachers apparently give a good deal of concentrated and highly verbal attention to reversal and rotation errors. But there is no evidence that teaching is required. It may be that the child learns which varying dimensions of letters are significant and which are not by simply looking repeatedly at many samples containing both varying and invariant features. The distinctive features of phonemes are not taught but they are nevertheless learned.

Helping the child to pay attention to the distinctive features can hardly hurt, however. If we knew the set of such features, they could be incorporated in some of the "reading readiness" tasks which involve visual discriminations. There is little or no evidence that these experiences transfer to reading. But if the typical matching tasks used variables which are significant for letter discrimination (instead of pictures of objects), there would certainly be greater potential transfer value.

SUMMARY AND CONCLUSIONS

This experiment studied qualitatively as well as quantitatively the development of visual discrimination of letter-like forms in children four through eight. The forms were constructed according to the same constraints which govern formation of printed capitals.

Twelve were chosen as standards, and 12 specified transformations were constructed for each standard. The transformations were three degrees of change from line to curve or vice versa; five changes in orientation; two perspective transformations; and two topological transformations. The discrimination task required S to match a standard form with an identical form. All Ss matched for all 12 standards. Errors were classified according to type of transformation erroneously identified with the standard. The experiment was repeated for the 5-year-old group using real letters.

Overall error scores decreased with age, but difficulty of discrimination was different for different transformations. Initial errors were greatest for perspective transformations and least for topological transformations, with changes of rotation and reversal in between. Changes of line to curve varied in difficulty depending on the number of changes.

The slopes of the curves differed, as well as the initial error. The four classes of transformation showed similarity of slope within the class, but significant differences between them. Errors made with real letters correlated significantly with errors made with letter-like forms. The correlation between mean errors for different transformations in the two sets of material was $+0.87$, showing that the kind of transformation defining similarity or difference between two forms is a good predictor of confusion errors.

The differences between the developmental error curves for the four types of transformation were interpreted in terms of a hypothesis of distinctive features. Features which have been in the past

critical for distinguishing objects are assumed to transfer to graphic discriminations. Discrimination learning continues from this point for distinctive features of letters, but proceeds slowly, if at all, for those varying features of graphemes which are not critical for distinguishing them.

REFERENCES

- Brown, R. Words and things. Glencoe, Ill.: Free press, 1958.
- Fitts, P. M., Weinstein, M., Rappaport, M., Anderson, N., & Leonard, J. A. Stimulus correlates of visual pattern recognition: A probability approach. J. exp. Psychol., 1956, 51, 1-11.
- Gibson, J. J., & Gibson, E. J. Perceptual learning: Differentiation or enrichment? Psychol. Rev., 1955, 62, 32-41.
- Gibson, J. J., & Gibson, E. J. Continuous perspective transformations and the perception of rigid motion. J. exp. Psychol., 1957, 54, 129-138.
- Jakobson, R., & Halle, M. Fundamentals of language. The Hague, Netherlands: Mouton, 1956.
- Klemmer, E. T. The perception of all patterns produced by a seven-line matrix. J. exp. Psychol., 1961, 61, 274-282.
- Piaget, J., & Inhelder, B. The child's conception of space. London, England: Humanities Press, 1956.
- Ryan, T. A. Significance tests for multiple comparison of proportions, variances, and other statistics. Psychol. Bull., 1960, 57, 318-328.
- Vernon, M. D. Backwardness in reading. Cambridge, England: Cambridge Univ. Press, 1957.

AN ANALYSIS OF CRITICAL FEATURES OF LETTERS,

TESTED BY A CONFUSION MATRIX

Eleanor Gibson, Harry Osser,
William Schiff and Jesse Smith

Cornell University

A developmental study of the discrimination of letter-like forms (Gibson, Gibson, Pick & Osser, 1962) showed that discrimination improved from 4 years of age to 8 years. But the rate of developmental change differed markedly depending on the transformation relationship between the forms which were confused. Differences between forms which could be described as topological transformations were discriminated quite well even at 4 years and confusions had disappeared by 8 years. Differences of perspective transformation, on the contrary, were seldom discriminated (80 percent errors) at 4 years and errors were still frequent at 8 (60 percent). Line-to-curve transformations and rotations were confused about 50 percent of the time at 4 years but errors decreased to 10 percent or less by 8 years. What factors are responsible for the change in discriminability over time and why are there qualitative differences depending on the transformation? In answer to this question it was suggested that what was learned during the 4-year interval was the distinctive features of letters (graphemes) and that discrimination had improved to near perfection with respect to those transformations which were critical for distinguishing graphemes from one another (e.g., topological, line-to-curve and rotational transformations) whereas it improved very little for the non-critical change (perspective transformation).

If this interpretation is correct, it would be extremely useful to know just what the "critical" or "distinctive features" of letters are. What dimensions of difference must a child learn to discriminate for each letter to become unique, never confused with another? One can make some fair intuitive guesses at these features, like the straight-curve and open-closed differences tested in the previous experiment. But can a list of features sufficient to account for grapheme differences be drawn up and put to some kind of test?

The method pursued for investigating this question was that of drawing up a "feature chart" for a given set of graphemes, obtaining a confusion matrix for these graphemes with pre-reading children, and comparing the errors in the matrix with those predicted by the feature chart. Predictions from the chart were based on the number of feature differences between the two graphemes--that is, the fewer the feature differences between two graphemes, the greater the number of confusions expected.

Programs for machine recognition of letters are relevant to this approach. Some of these programs are "prototype" methods, that is matching to a template without analysis, and are of little interest to us. Some, however, are based on a "property list" comparable to the set of features contemplated in our approach (see, for example, Selfridge & Neisser, 1960; Uhr, 1963). However, the objective is different; our purpose was not to achieve an economical set of operations for a machine, but to get leads toward what it is that a human child is learning as grapheme discrimination develops. What features or dimensions of letter difference do actually serve to distinguish them for the child, or lacking them, make them confusable?

The method outlined was greatly influenced by Jakobson and Halle's analysis of the distinctive features of phonemes (Jakobson & Halle, 1956). A table of 12 or so features, each in binary opposition, serves to give a unique pattern for all phonemes, so that any one is distinguishable from any other by its pattern of attributes. A pair of phonemes may differ by any number of features, the minimal distinction being one feature opposition. The features must be invariant under certain transformations and essentially relational (a contrast), so as to remain distinctive over a wide range of speakers, intonations, etc. Brown and Hildum (1956) argued that if phonemes are distinguished from one another by differences in their feature patterns, it should follow that the smaller the number of feature oppositions between two phonemes, the greater should be their confusability under conditions of impoverishment or noise. They tested this prediction by playing a recording of syllable triplets and having Ss write what they heard. The results were analysed for mistakes involving a change in a single phoneme and these mistakes were classified according to the number of significant feature changes compared to the syllables which had been recorded. Errors did in fact tend to be associated with minimal feature changes. A similar test of grapheme confusability was our purpose, except that we wished to obtain the data from children, under reasonably natural, noise-free conditions.

THE FEATURE LIST

How should a list of the critical features serving to distinguish graphemes be chosen? Human intuition so far is the only "generator," but intuition can be assisted by considerable evidence in the literature

of developmental and physiological psychology. In the course of several years' study of this problem, a number of feature lists have been drawn up, tried against the available criteria, and revised. The immediate criteria, prior to the results of the matrix, were as follows:

(1) features must be critical ones, present in some members of the set but not others so as to present a contrast; (2) they must be invariant under perspective and size transformations; (3) they must yield a unique pattern for each grapheme; (4) the list must be reasonably economical. If the list contained 26 features, we should have done no better than a template matching scheme. On the other hand, cue redundancy is characteristic of human ecology and psychology, so economy is not the primary criterion.

Our earlier experiment suggested a beginning. The straight-curve distinction could be critical, since it exists as a contrast between letters, and precision of its discrimination increases developmentally; likewise the open-closed distinction. The rotational one is clearly important, but is hard to specify in a feature list (computer programs for recognition often make use of N, S, E, W, NW, etc., for directional analysis, but even with great proliferation of directions, the application is untidy). In the end, two feature lists were tried out against results of the confusion matrix. Each of them gave a unique pattern for the 26 letters of the type chosen for test. These letters were a simple Roman alphabet, capitals, available on a sign-typewriter with large characters, and they were used to obtain the confusion matrix as well.

Feature List I

1. Straight

a. Direction

Horizontal

Vertical

Oblique

Oblique

b. Section

Bisects

Intersects

2. Curve

a. Closed

b. Open vertically

c. Open horizontally

3. Number of strokes

1

2

3

4

4. Symmetry

Each feature in the list above is or is not a characteristic of each of the 26 graphemes. For instance, one asks the question, "Is there a curved segment?" and gets a yes or no answer. For an A, for example, there is no curve. But there are yes answers for straight, horizontal, and both obliques. There is a yes for section (bisection), for three strokes, and for symmetry. A filled-in feature chart shows the pattern of yes answers for each letter. Each pattern is unique. The "decision" for each letter can proceed to some extent in an orderly progression, but there is not a perfect hierarchy like Jakobson's for the phoneme features. A hierarchy with progressive splitting probably does not exist for distinguishing graphemes. The number of potential features is very large, and varies considerably from one alphabet and one type font to another.

The features in the list above were picked with some rationale or presumed heuristic value in mind. Consider, for instance, the curved-straight distinction. Evidence from neurophysiological research with frogs (Maturana, Lettvin, McCulloch & Pitts, 1960) found that straight edges and convexities were responded to as units. Other evidence with cats (Hubel & Wiesel, 1962) suggests that orientation of a line (e.g., horizontal, vertical, oblique) is responded to uniquely by specific single fibers. Evidence from work on stabilized images with man (Pritchard, Heron, & Hebb, 1960) suggests that straight lines and curves disappear as units, breaks occurring at points of intersection. Thus, straight lines, curved lines, horizontals, verticals and obliques may have a kind of elementary status.

The closed-open distinction for curves was suggested by Piaget & Inhelder's (1956) developmental study showing that topological differences are ontogenetically primitive ones. Intersection is another topological characteristic which, according to Piaget, can be distinguished very early.

Number of strokes was included in this list as an indicator of relative complexity, and symmetry as an indicator of redundancy.

The second feature list which was tested also included the curved-straight distinction. Number of strokes was discarded. Intersection was not split into two types. Two other features were added, a scanning feature and cyclic change (another index of redundancy). The scanning feature is split into two; a discontinuity following a vertical scan (scan downward), and a discontinuity following a horizontal scan (scan left to right). Directional scanning habits are required by every writing system and are reflected in the habits of a

Feature List II

1. Straight
 - Vertical
 - Horizontal
 - Oblique
 - Oblique
2. Curved
 - Closed
 - Open vertically
 - Open horizontally
3. Intersection
4. Redundancy
 - Cyclic change
 - Symmetry
5. Discontinuity
 - Vertical scan
 - Horizontal scan

skilled reader (Heron, 1957, and others). Theories of shape discrimination by means of analysis in terms of horizontal and vertical scans have received considerable attention recently (see, for instance, Sutherland, 1963). Discontinuity, which stops the scan, is itself a topological feature which seemed to have some intuitive "rightness."

The cyclic change feature was suggested by the fixed image work of Pritchard, Heron and Hebb (1960) which showed that parallels act as units in appearing and disappearing.

This feature list combines three types of features: elements, features dependent on habits required by the writing system, and what might be termed "Gestalt" features (intersection, symmetry, cyclic change). The mixture is "impure" and therefore not as elegant as a homogeneous list, containing only the same type of unit of analysis, but all have been shown to affect discriminability and are

embodied in the writing system. The question is, are they critical? The best test we could conceive was to compare extent of feature differences for pairs of letters with the confusions which children actually make between them.

THE CONFUSION MATRIX

Method.

Since we wanted an error matrix from children who did not yet read, a judgment easy enough to be comprehended by four-year olds was sought. Learning to identify letters by name was discarded for several reasons, one being that most children usually knew the names of a few letters (e.g., first letters of their own names) and showed a strong response bias toward them.¹ A same-different judgment, without impoverishment of the stimuli, yielded almost no errors. A matching method was therefore worked out, with one letter presented as the standard and a multiple choice set of six letters (one the standard) presented immediately afterward for S's matching judgment. The letters in the choice group were randomly selected, but the sets of choices were arranged so that every letter appeared an equal number of times in the choice sets for every standard. Furthermore, the sets of letters were prepared so that every standard appeared once in all six positions in order to balance out possible tendencies to choose on a position basis. Order of presentation of standards was randomized, with order balanced over a number of lists. Thus every letter had an equal opportunity to be mistaken for any other, without bias from order effects.

1. Even confusion errors obtained from adults using an identification response for letters set at a distance (in the manner of a Snellen chart) showed strong effects of "response bias" (habits of responding which produced individual peculiarities). (In Sanford, 1887.)

The material was typed with a sign-typewriter in Roman capitals 1/4 inch high on strips of adding machine paper which fit a small memory drum (see Figure 1). The memory drum was used to present the material. It was timed for a 1-second presentation of the standard letter, followed by a 3-second presentation of the choice set. A 4-second empty interval passed before the next standard appeared.

Figure 1. Sample of List

 K
K S A W Y O

 Y
E Q V Y W H

 G
R G T Q J B

 I
U J N B D I

The S held a wooden stylus with which he indicated the letter in the choice set which he thought matched the standard he had just seen. His instructions were given slowly and repeated as often as necessary to make certain he understood the task. Then some practice was given with numbers (arranged in the same fashion as the letters) to accustom the child to the timing and to check on his comprehension of the instructions. If a child seemed incapable of performing the task and attending to the stimuli at this stage, he was dropped, but this was seldom necessary. If the child's attention flagged during the task, the drum was stopped and the experimenter attempted to recall him to the task.

Since balancing for order effects and equating opportunities for confusion made a great many comparisons necessary, no four-year-old child could go through more than a small fraction of them. A balanced block design was therefore adopted. It required 180 blocks of 26 judgments to be completed. Most of the children went through only one block of judgments. The slightly older or more attentive children went through two blocks. Going through one block, every child matched every letter of the alphabet as standard with one of the choice sets.

Since our only information came from errors, and since there was very little impoverishment in viewing conditions, it was decided to replace a child who made absolutely no errors with another child in that block of the design. Because of these replacements, and the children who could not be run due to lack of comprehension, a very large number of subjects was required--all the four-year olds and a few slightly older and younger children from five nursery schools.² The 180 blocks of data were obtained from 87 children.

Results.

The confusion matrix obtained is presented in Table 1. Since the two halves were quite symmetrical, they have been combined. The numbers in most cells are zero or very small, even so. Despite the large investment of experimental time, the yield of errors was minimal, due to careful instruction and urging to attend.

2. We wish to thank the staff and children of the following schools for their interest and cooperation: the Cornell University Nursery School, Henry St. John's Nursery School, The West Side House, Lent's Nursery School, and the Country Play School.

[illegible]

Because the number of errors was not large, it was decided that a replication was advisable. Some changes in procedure were required, however, because the available subject supply was nearly exhausted.

REPLICATION OF MATRIX

Method.

A more economical method was devised, with the following changes. First, instead of rotating letters through every order in the choice sets, random orders were employed. Position of a standard in a list was randomized as well, and the number of list orders reduced to four.

Second, two sets of lists were devised, one to be a "high confusion" list and one to be a "low confusion" list. The purpose of this modification was dual. The high confusion list, if our predictions were correct, should have a much higher rate of errors than the low confusion list. This would have the advantage of giving us more data and also the possibility of validating predictions from our first matrix and our feature list.

The high confusion list was constructed by selecting for each letter the most frequent confusions which had occurred in the matrix. If there were only two or three letters which had been confused, the other letters included in the set of six were selected according to the feature list, those letters with minimal feature differences being included. For the low confusion list, the set for each letter included only letters which had not been confused at all in the matrix, and which had maximal feature differences.

The procedure for obtaining the judgments was also modified slightly. Instead of running the drum at a fixed speed, the child was allowed to take as long as he needed to respond, and his reaction time was recorded

by means of an electric key-operated pen on a continuously moving paper roll (Esterline Angus). Each child was run through both a high and a low confusion list, making 52 judgments.

Subjects.

The 20 Ss were again obtained from a nursery school. None had been included in the first group.

Results.

The results are presented in the confusion matrix in Table 2. Since this matrix is not cell by cell comparable with the first one (some letters having been omitted from some of the choice lists), our best evaluation is a comparison of the lists which were predicted to have high and low confusability. Results for the two lists can be compared with respect to errors and also with respect to reaction time. Many more errors were made in the list predicted to have high confusability (99, compared with 14 in the other list, $p < .001$ by t test, two tailed). None of the 20 Ss reversed the trend. The four subgroups which had different list orders showed nearly identical differences in the same direction. The times differed significantly as well. Decision times were longer for judgments in the high confusion series ($p < .001$).

The results in general, therefore, tend to confirm the major differences in the first matrix (high and low confusion letters) and to increase our confidence in the feature lists.

TABLE 2

Confusion Matrix Experiment II

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	2
B		0	0	1	0	2	1	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
C			0	0	0	7	1	0	0	0	0	0	1	1	0	0	0	1	0	2	0	0	0	0	0	1
D				0	0	0	0	0	0	1	0	0	0	3	0	2	0	0	0	0	0	0	0	0	0	1
E					4	0	4	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
F						0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
G							0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
H								0	0	1	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0
I									0	0	6	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
J										0	0	0	0	0	0	0	1	0	1	0	2	0	0	0	0	0
K											0	0	0	0	0	0	0	0	0	0	0	1	0	4	3	0
L												0	0	0	0	0	0	0	0	4	0	0	0	0	0	2
M													7	0	0	0	0	0	0	0	0	0	3	0	0	0
N														0	0	0	0	0	0	0	0	0	3	0	0	0
O															0	0	0	0	0	0	0	0	0	0	0	0
P																2	0	0	0	0	2	0	0	0	1	0
Q																	5	0	0	0	0	0	1	0	0	0
R																		0	0	0	0	0	0	0	0	0
S																			0	0	0	0	0	0	0	0
T																				0	0	0	0	0	0	0
U																					2	0	0	0	2	0
V																						1	0	0	0	0
W																							0	0	5	0
X																								0	0	0
Y																									2	0
Z																										2

Too much cannot be hoped from this rather untidy procedure, but if the feature lists are predicting anything, it might still be expected to show up in some significant, if low, correlations. To provide a yardstick for evaluating them, it was decided to correlate an overlap measure with the errors as well. The overlap measures were taken from a base-Warbridge report (1961). They were obtained by placing one letter over another and measuring the amount of light transmitted. Such measures were obtained for several kinds of type, one of them fortuitously very similar to the one we had used in obtaining the matrix. The overlap matrix obtained with this type was correlated with our error matrix.

The three sets of correlations are presented in Table 3. They are not high, and a number are insignificant, but in all three sets the

PREDICTION OF ERRORS BY FEATURE DIFFERENCES

To say that our confidence in the feature analysis has increased is a rather loose statement. Some quantitative test of whether the feature lists are really predicting anything in the error matrix is desirable. Correlations were consequently run between percent feature differences and number of confusions, one correlation for each letter of the alphabet. To calculate percent difference in feature patterns, the pattern of features of a given letter was compared with another; the number of plus features entered for one letter which were not marked identically for the other were divided by the total number of pluses for the pair. This is admittedly a crude system, since it is quite likely that some features should be weighted more heavily than others. Since confusion errors for a given pair were often zero, and our quantities not too precise, a Spearman rank order method was chosen. Also, the two error matrices were combined to increase the number of errors in the cells and the overall variance.

Too much cannot be hoped from this rather untidy procedure, but if the feature lists are predicting anything, it might still be expected to show up in some significant, if low, correlations. To provide a yardstick for evaluating them, it was decided to correlate an overlap measure with the errors as well. The overlap measures were taken from a Ramo-Wooldrige report (1961). They were obtained by placing one letter over another and measuring the amount of light transmitted. Such measures were obtained for several kinds of type, one of them fortunately very similar to the one we had used in obtaining the matrix. The overlap matrix obtained with this type was correlated with our error matrix.

The three sets of correlations are presented in Table 3. They are not high, and a number are insignificant, but in all three sets the

Table 3

Correlations between Confusion Errors and Three Predictive Measures

Measure	Feature List 2	Overlap	DF 2
A	.274	.037	.212
B	.038	.374	.092
C	.427	.583	.434
D	.105	.336	.213
E	.223	.146	.280
F	.339	.449	.429
G	.409	.560	.547
H	.446	.291	.401
I	-.059	.135	.088
J	.505	.229	.614
K	.496	.539	.491
L	.234	-.104	.303
M	.422	.300	.396
N	.131	.389	.183
O	.402	.342	.555
P	.424	.222	.529
Q	.117	.398	.223
R	.325	.370	.304
Sq	.183	.233	.151
T	.436	.030	.334
U	.299	-.048	.278
V	.387	-.240	.435
W	.499	.280	.491
X	.447	.480	.503
Y	.103	.425	-.180
Z	.022	-.202	-.043

number of significant correlations is far greater than chance. There are 12 significant r 's for feature list II, 11 for feature list I and 8 for the overlap measure. Feature list II thus looks best, but it is not clear how much confidence should be placed in its superiority. Since the overlap measure has four negative r 's, as well as fewer significant positive ones, the feature analysis seems to be doing better.

These correlations afford, in a way, a comparison between a template-matching model and a property-analysis model of letter-discrimination. The overlap measure is analogous to a template match and the feature-difference measure is a property analysis. So far as we know, none of the machine recognition programs have as yet been tested for confusions predicted by the model.³

Predictions from feature differences to errors are considerably more successful for some letters than for others. It may be possible by studying these to refine the feature list and introduce some weights. However, the very low correlations tend to be associated with a low error rate for all the pairs and little variance. It would be desirable, of course, to obtain more data with a method which produces a higher error rate, but we have so far been unable to devise a way of doing this without introducing noise or impoverishment in the stimulus conditions.

Another likely explanation for the lowness of some of the correlations is that our Ss were children who had not yet learned the alphabet and were not, consequently, able to select out all of the distinctive

3. Dr. Frank Rosenblatt of the Cornell University Cognitive Systems Research Program has been given our confusion matrix for comparison with discrimination learning times on the "perceptron." He predicted that time required for the perceptron to discriminate a given pair of letters should be related to the frequency of confusions in the matrix.

features. They might have failed to discriminate where a contrast exists which appears obvious to an adult but is nevertheless a late one to be selected out. The redundancy features, symmetry and cyclic change, might be of this kind, whereas a straight-curve contrast might be picked up earlier. An error matrix secured from adults would be useful for comparison, but it would have to be obtained under different conditions.⁴

MULTI-DIMENSIONAL ANALYSIS OF THE MATRIX

Torgerson (1958) has a computer program for analyzing a confusion matrix, considering the matrix as a set of similarity relationships (proximities) representing distances in a multidimensional space. An analysis carried out by Dr. Torgerson⁵ resulted in a three-dimensional solution, two dimensions of which were readily identifiable as 1) straight vs. curve, and 2) relative obliqueness. The third dimension might conceivably be thought of as complexity. A study of the plots representing this dimension suggests that several modes of difference may be mixed in it. It is interesting that the closed letters cluster in a small area, with the open characters at considerable distance from them toward the perimeter. The small numbers in the cells and the frequency of empty cells make analysis into further dimensions impossible, but it is encouraging that dimensions 1 and 2 correspond with features prominent

4. A small amount of data was obtained from 5 adults by Korte (1926) using peripheral exposure and an identification response. Letters confused by more than one S were C and G, C and O, D and B, E and F, E and B, F and P, G and O, H and R, I and L, P and R, P and B, Q and G, R and B, X and K, X and Y, Y and V. The data suggest that a closed-open contrast disappears very quickly with this method.

5. We are deeply grateful to Dr. Torgerson for his interest and for carrying out this analysis for us.

in our analysis. Perhaps these two dimensions have priority genetically and in the discrimination process, and should be heavily weighted in a feature analysis.

CONCLUSION

An analysis of the differences between graphemes of a specified type was made in terms of potential distinctive features of graphemes. Two feature lists, both of which gave a unique pattern of features for each of the 26 letters, were used to predict errors in a confusion matrix obtained with the same letters. The subjects (four-year-old children) performed a matching task with the letters. For both feature lists, the percent feature difference between letters was correlated with the errors. An overlap measure was also correlated with the errors. The number of significant correlations was far greater than chance for all three measures, the second feature list being the best predictor. The choice of certain feature dimensions (curve-straight and obliqueness) was corroborated by a multi-dimensional analysis of the matrix.

REFERENCES

- Brown, R. W., and Hildum, D. C. Expectancy and the perception of syllables. Language, 1956, 32, 411-419.
- Gibson, E. J., Gibson, J. J., Pick, A. D., and Osser, H. A developmental study of the discrimination of letter-like forms. J. comp. & physiol. Psychol., 1962, 56, 897-906.
- Heron, W. Perception as a function of retinal locus and attention. Amer. J. Psychol., 1957, 70, 38-48.
- Hubel, D. H., and Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol., 1926, 160, 106-154.
- Jakobson, R., and Halle, M. Fundamentals of language. The Hague, Netherlands: Mouton, 1956.

- Korte, W. Über die Gestalt-auffassung im indirekten Sehen. Zeits. f. Psychol. und Physiol. der Sinnesorgane, 1923, 93, 17-82.
- Maturana, H. R., Lettvin, J. Y., McCulloch, W. S., and Pitts, W. H. Anatomy and physiology of vision in the frog (*Rana Pipiens*). J. gen. Physiol., 1960, 43, 129-175.
- Piaget, J., and Inhelder, B. The child's conception of space. New York: Humanities Press, 1956.
- Pritchard, R. M., Heron, W., and Hebb, D. O. Visual perception approached by the method of stabilized images. Canad. J. Psychol., 1960, 14, 67-77.
- Ramo-Wooldridge. Technical Report, Part II, on Word Searching Device, C 117- 1U14. Canoga Park, Calif., 1961.
- Sanford, E. C. Relative legibility of the small letters. Amer. J. Psychol., 1, 1887, 403-435.
- Selfridge, O. G., and Neisser, U. Pattern recognition by machine. Scient. Amer., 1960, 203, 60-68.
- Sutherland, N. S. Shape discrimination and receptive fields. Nature, 1963, 197, 118-122.
- Torgerson, W. S. Theory and Methods of Scaling. New York: Wiley, 1958.
- Uhr, L. "Pattern recognition" computers as models for form perception. Psychol. Bull., 1963, 60, 40-73.

IMPROVEMENT IN VISUAL DISCRIMINATION OF LETTER-LIKE FORMS¹

Anne D. Pick

Gibson, Gibson, Pick and Ouser(1962) in a developmental study of the discrimination of letter-like forms, found that the discriminability of these forms increased for subjects between the ages of 4 and 8. That is, errors in a matching task decreased as the age of the subjects increased.

The question of how this increase in discriminability of graphic symbols occurs has obvious implications for education as well as theoretical importance. People have asked such questions as these: Is a discrimination, once learned, specific to the forms on which it was practiced, or can it transfer to other forms? What is the process or mechanism involved in learning a discrimination? What is the function of practice in increasing discriminability? These problems have stimulated hypotheses about what is learned during improvement of discrimination and what is the basis for transfer of a learned discrimination.

One hypothesis (see, for instance, Vernon, 1952, 1955, 1959, and Bruner, 1957, 1957b) is that discrimination and identification involve matching sensory data or "cues" about objects to schemata or prototypes of the objects which have been built up through repeated experience with the objects, and which are "stored" in memory. Improvement in discrimination would involve first constructing schemata of the objects to be discriminated, and then matching the

1. This paper is a summary of a dissertation presented in partial fulfillment of the requirements for the Ph.D. degree under the direction of E. J. Gibson.

sensory data to the schema to identify it as "same" or "different." Practice with objects to be discriminated would enable the subject to build up and refine the appropriate schemata. Bruner suggests that "discriminatory cues" are involved in the matching process, and by this he means that "the properties of stimulus inputs make it possible to sort these inputs into categories of best fit" (1957b, pp. 132-3). In other words, perceptual discrimination occurs when a percept is constructed from sensory "cues" and this percept is matched to an existing schema or category. The discriminatory cues serve to locate a stimulus input in a specific category. If the category is a narrowly defined one, then a fine discrimination is made; if it is a larger, more inclusive category, then a cruder discrimination is made.

According to this hypothesis, then, the way in which graphic symbols increase in discriminability would be by way of schema formation and perfection. A schema for the letter would be constructed and further delimited, and the immediate stimulus input would be matched against existing schemata.

Another hypothesis proposed by E. J. Gibson (1962) about the nature of improvement in discrimination is that it involves the detection of distinctive features. Distinctive features are differences between objects rather than prototypical models of the objects.

Improvement in discrimination would consist of learning to perceive or respond to variables of stimulation which distinguish between the objects to be discriminated. These distinguishing features are dimensional variables which provide contrasts between objects. The function of practice, according to this point of view, is to enable

the subject to respond to an increasing number of stimulus variables, and to enable him to discover which of these variables are "critical" in the sense that they serve to distinguish between one object and another. Discrimination, then, consists of responding on the basis of these critical differences.

These two hypotheses differ in what is postulated to be learned during improvement in discrimination. For one, specific schemata of the objects to be discriminated are presumably developed. For the other, the general stimulus variables corresponding to the distinguishing features of the objects are detected.

The physical conditions necessary to produce improvement in discrimination do not necessarily differ for the two theories. Both would predict that such improvement will occur as a function of practice with the objects to be discriminated. Neither postulates the necessity of learning verbal responses to the objects. The same learning conditions would allow for the construction of a schema or for the detection of critical features. A transfer experiment is necessary to determine the extent to which prototype learning and detection of distinctive features have actually occurred during training.

With a transfer design, subjects can be presented with initially undifferentiated stimuli and trained to discriminate between them. They can then be asked to discriminate between new stimuli which either differ from each other in the same dimensions as those which they learned to discriminate or which have the same prototype or standard as those which they learned to discriminate. Differential performance in these two test situations should help to decide between the two

theories discussed. Better performance with new stimuli having the same standards as those used in training would suggest that schemata had been constructed. Conversely, better performance with new stimuli varying in the same dimensions as those used in training (but different standards) would suggest that distinctive features had been detected.

The present experiment is designed to explore this problem with respect to graphic forms like those used by Gibson, Gibson, Pick, and Osser (1962).

METHOD

Subjects .

The subjects were 60 kindergarteners from South Hill School randomly assigned to each of three equal groups ($n = 20$ in each group). There were an approximately equal number of boys and girls in each group.

Materials.

The stimuli were letter-like forms (used in the earlier developmental study) constructed according to the rules which govern the construction of real letters. The forms were photographed and mounted on white cardboard cards about 2" x 3". There were six standard forms and six transformations of each of the standard forms. These transformations were two changes of line to curve, a right-left reversal, a 45^0 rotation, a perspective transformation, and an increase in size.² A form board, similar to a reading stand was used to

2. These transformations are not necessarily those which provide dimensions of contrast or distinctive features for real letters. A size transformation, for example, is not a difference which is critical for letter discrimination, but it was made critical here for experimental purposes.

display the standard forms in front of the subject. The back of this form board was inclined approximately at a right angle to the subject's line of sight.

Procedure.

TRAINING: The training procedure was the same for subjects in all three groups. Three standard forms were placed on the board in front of the subject. Then a pack of cards containing two copies of each standard and three transformations of each standard was spread out on the table between the S and the board containing the standards. S was instructed to go through the pack, looking carefully at each card, and decide whether it was exactly the same as one of the standards, or if it was different. When S found one that was exactly the same as a standard, he gave it to the experimenter. When S had completed a trial, i.e., when he had made a judgment about every card in the pack, the cards were shuffled by E, and the procedure was repeated until S completed one perfect trial. Matching errors and trials to criterion were recorded. A correction procedure was followed on every trial except the first. A practice trial with real letters and correction was carried out before the start of the experiment. This practice trial was included to acquaint S with the procedure and to ensure that he understood that only forms which were "exactly the same" as the standard should be given to E. One S was omitted from the experiment on the basis of this practice trial because she did not understand the concepts of "same" and "different." Another S was substituted so that the number of subjects in each group would remain equal.

TRANSFER: Following one perfect training trial, the transfer procedure was carried out. There were three transfer groups, and the one for a given subject was predetermined. The task was presented as in training and was the same for all three groups except for the particular forms used. One trial without correction was given, and matching errors were recorded as before.

Group C, the control group, received three different standards and three different transformations of these standards from the ones used in training. This group was included to provide a baseline from which to measure the amount of transfer shown by the experimental groups. It would be expected that Ss in this control group would show some improvement in transfer as compared to their first training trial as a result of "warm-up," familiarity with the task, etc. It was therefore deemed advisable to compare the transfer performance of the experimental groups with the control group rather than with initial performance in training.

Group EI received the same standards, but different transformations from the ones used in training. It was reasoned that if what was learned during training were schemata of the standards, and if a schema is necessary for discrimination, then this experimental group should show more transfer than the other experimental group and the control group since they were dealing with the same standards as they had in training.

Group EII received different standards, but the same transformations of these standards as the ones used in training. In other words, these subjects were dealing with new forms, but the differences

between the forms were on the same dimension as the differences between the forms with which they had dealt during training. It was reasoned that if, during training, the subjects were learning to detect the differences between the forms, then this group should show more transfer than the other experimental group and the control group since they were dealing with the same kinds of transformations as they had in training.

There were four subgroups in the training condition, each of which used a different combination of standards and transformations. There were also four subgroups within each of the three transfer groups since the combination of forms used in transfer for a given subject depended on the combination of forms used in training. These different combinations of forms were introduced in order to control for the possibility of some forms being initially more discriminable than others.

RESULTS

Training

Table 1 shows the data for the training trials for the three groups.

Table 1. Initial Errors and Trials to Learn

Group	N	Confusion Errors in First Trial			Trials to Criterion (not including Criterion Trial)		
		Total	Mean	SD	Total	Mean	SD
EI	20	146	7.30	1.95	67	3.35	1.15
EII	20	151	7.55	2.01	69	3.45	.97
C	20	140	7.00	1.95	64	3.20	1.08

T tests were used to analyze the differences between the three groups, in number of confusion errors (incorrect matches) made on the first trial and in number of trials to the criterion of one perfect trial.

On neither of these measures did any difference approach the .05 level of significance. Therefore, any differences found between the three groups in number of errors made on the transfer trail can not be accounted for by differences in initial ability or by differences in amount of training received.

Transfer

Confusion errors in the transfer trial constitute the main data of the experiment. Table 2 shows the errors made in the transfer trial by each of the 12 subgroups. These subgroups represent the three transfer groups and the four combinations of standards and transformations used in each of these groups.

Table 2. Confusion Errors in Transfer Trial for the Twelve Subgroups (column headings indicate the particular combination of standards and transformations used in the transfer trial. N = 5 in each subgroup).

Standards Transformations Group	Sabc				Sdef			
	T123		T456		T123		T456	
	Total	Mean	Total	Mean	Total	Mean	Total	Mean
EI	12	2.4	17	3.4	21	4.2	19	3.8
EII	11	2.2	10	2.0	6	1.2	12	2.4
C	23	4.6	32	6.4	21	4.2	25	5.0

An analysis of variance was performed on the data for confusion errors for the 12 subgroups and a summary of this analysis is shown in Table 3. The effect of transfer groups is significant. The probability of obtaining an F as great as the one obtained (12.69) by chance is less than .001. There was no significant effect of either transformations or standards and none of the interactions was significant.

Table 3. Summary of Analysis of Variance of Confusion Errors Made in Transfer Trial by Subgroups

Source	SS	df	ms	F	p
Total	305.00	59			
Transfer groups (3)	96.15	2	48.08	12.69	<.001
Standards (2 combinations)	.03	1	.03	-----	
Transformations (2 combs.)	7.37	1	7.37	1.93	>.05
Groups X Standards	10.52	2	5.26	1.39	>.05
Groups X Transformations	2.78	2	1.39	-----	
Transformations X Standards	.40	1	.40	-----	
Groups X Stand. X Trans.	5.75	2	2.88	-----	
Error	182.00	48	3.79		

This indicates that the differences between the three experimental groups hold regardless of subgroups, i.e., regardless of the particular combination of standards and transformations used.

Table 4 shows the overall means (subgroups combined) for the three transfer groups. T tests were used to analyze the significance of the differences in transfer errors between the three groups. All three groups were significantly different from each other with a probability (two-tailed) of less than .01.

Table 4. Total and Mean Number of Errors Made in the Transfer Trial by Three Transfer Groups (subgroups combined. N = 20 in each group)

Group	Total Errors	Mean Errors
EI	69	3.45
EII	39	1.95
C	101	5.05

DISCUSSION AND CONCLUSIONS

The results imply the importance of learning dimensions of difference in improvement of visual discrimination. Subjects who were dealing with forms which they had never seen before but which differed from each other in familiar ways made fewer errors than subjects who were dealing with standards with which they had had experience. The fact that Group EI, the group having familiar standards and new transformations, was superior to the control group, suggests that the subjects constructed schemata during training. However, the clear superiority of Group EII (the group having new forms but the same transformations as in training) over Group EI implies that schema formation may not be essential to discrimination learning of this sort, and that learning to pick out dimensions of difference may be a much more efficient way to learn discriminations such as those required in this experiment.

It is concluded that learning dimensions of difference which provide contrasts (distinctive features) is a significant component of improvement in visual discrimination of forms such as letters.

Subsequent experiments have explored the generality of these results for improvement in tactual discrimination.

REFERENCES

- Bruner, J. S. Neural mechanisms in perception. Psychol. Rev., 1957, 64, 340-358.
- Bruner, J. S. On perceptual readiness. Psychol. Rev., 1957b, 64, 123-152.
- Gibson, E. J., Gibson, J., Pick, A.D., and Osser, H. A developmental study of the discrimination of letter-like forms. J. comp. physiol. Psychol., 1962, 55, 897-906.
- Vernon, M. D. A Further Study of Visual Perception. Cambridge, England: Univ. Press, 1952.
- Vernon, M. D. The functions of schemata in perceiving. Psychol. Rev., 1955, 62, 180-192.
- Vernon, M. D. Perception, attention and consciousness. The Advancement of Science, 1959, 62, 111-123.

A POSSIBLE EFFECT OF LEARNING TO WRITE

ON LEARNING TO READ

James J. Gibson and Harry Osser
Cornell Univeristy

In the acquisition of language it is clear that a child never learns to understand speech without at the same time learning to speak. The circularity of speaking and hearing has always been recognized and the degree to which the auditory feedback controls the process of uttering words has recently been studied experimentally. What has not been so clear is that a child never learns to read without at the same time learning to write. Reading and writing are different terms and often seem to be thought of as different school subjects. Nevertheless, the visual feedback which controls the manual art of writing is quite analogous to the auditory feedback which controls the vocal art of speaking. Just as one cannot speak without hearing one's speech, so one cannot write without, in a peculiar sense, reading one's writing. The flow of stimulation is self-produced instead of other-produced, but the conventional identities between the stimuli made by one self and those made by others are the basis for symbolic communication.

Writing is, of course, a much more complex psychological achievement than speaking, and this is evident both in the development of the child and in the evolution of the human species. Children begin vocal communication at about one year of age, but they cannot begin graphic communication until much later--certainly not until the phoneme elements of their speech are sufficiently

crystallized, as it were, to enable them to learn the remarkable fact that graphic elements correspond to them. What the child must learn, in order to read and write, is that speech sounds can appear, and can be made to appear, on a visible surface.

The act of making marks on a surface is something that a young child does spontaneously even when he is not given special opportunities to do so. Making a trace and looking at it is fascinating, even for monkeys. It might be called the fundamental graphic act. The leaving of a visible change on a previously blank surface--a trace, scratch, impression, line, or deposit of pigment--is something that is bound to catch the attention of a primate, and also to educate his hand, for a trace is an exact record of the movement of the hand. The fact that the trace continues after the movement has ceased, however dimly perceived, is surely an exciter of curiosity and the basis of further learning.

The fundamental graphic act is performed by children in the form of scribbling or finger-painting or mud-dabbing long before they begin to make representations or drawings. Just so, presumably, our hominid ancestors scribbled on the impressionable surfaces of their caves long before the human cave-painters discovered the possibility of representation. The making of a mammoth on a cave wall was an act (and the subsequent seeing of the mammoth was a percept) of a sort that must have inspired awe. This discovery, once made, was practiced by men for many thousands years and surely contributed to the development of thinking.

The representing of speech, of words as contrasted with visible objects, was a still more difficult and sophisticated graphic act,

and its discovery was made only a few thousand years ago. Writing-reading is said to mark the beginning of civilization and, in the case of the child, it marks the beginning of formal education. Literacy was difficult for man to achieve, and it is not easy for the child to learn. Unlike speech, it has to be taught. How best to teach it is a controversial question of psychology. The question is usually formulated, however, in terms of the teaching of reading, and the possible effect of learning to make letter sequences on learning to interpret letter sequences is left out of account.

THE HYPOTHESIS

In a phonetic writing-system (as distinguished from an ideographic writing-system like Chinese) there is a cumulative display of graphic elements in adjacent order which correspond with a set of phonemic elements in sequential order. In European systems the adjacent order is from left to right and from top to bottom of the surface on which the graphemes are displayed. This is convenient for direct manual production of a letter-sequence, handwriting, and is conventional for machine-mediated production, or writing in typed letters. However sophisticated the act of writing may become, however, it is always fundamentally the making of a visible mark, trace, pattern, or form on a surface, and leaving it there. It is always a graphic act as defined above.

The question arises, therefore, whether learning to read is quite as independent of learning to write as the teaching procedures used in schools often assume. The education of the child's

attention to the special features of and special differences between graphic symbols is necessary for reading. This might well be facilitated by letting the child make and look at these various symbols instead of just looking at them. The child has been causing marks to appear on the surface for years and it is time for him to begin making alphabetical marks. Perhaps this is the best way to train him to discriminate them.

The hypothesis is that practice in letter-making will facilitate letter-recognition. It is assumed that the essence of letter-making is the leaving of a visible mark specific to the motor act. If the motor act is performed but the specific visible mark does not appear (as in using a pencil with no lead, or a typewriter with no ribbon) then the graphic act is frustrated and this is not writing as defined.

For testing this hypothesis, an experimental group of children was given a pretest and a post-test of letter-recognition, with intervening practice in making letters by pressing the keys of a typewriter. A control group was treated in exactly the same way and performed the same motor reactions but without leaving traces. These children could feel and see the action of the type-bar, like the others, but the surface remained blank.

METHOD

Subjects

Twenty children from 4- to 6-years-old from day camps run by the Ithaca Board of Education served as subjects.

Each response was recorded as it was made.

Apparatus

The main piece of apparatus was a Smith-Corona Standard Secretarial typewriter with type 3/16 inch high. Certain alterations were made to the typewriter: (1) the line indicators were removed so that a clearer view of the letter was provided, (2) all non-letter keys were covered up with small pieces of adhesive paper, (3) the shift lock was set down so that only the upper case letters could be typed. In addition, the typewriter could be, at will, made not to "write" in the sense defined above, although its mechanical action remained unaltered.

Further equipment included a stopwatch and a set of 13 2x2 cards on which were typed the following upper case letters: A, D, G, I, J, L, N, P, R, V, W, X, Y. These were the alternate letters from each of the three rows of the typewriter keyboard.

Procedure

The first subject to arrive in any one experimental session was typically placed in one group (experimental or control) and the next to arrive in the other group. The final distribution of ages was two 4 year-olds, seven 5 year-olds, and one 6 year-old in each group.

Pretest. All Ss were given a pretest which consisted of an alphabet recognition test of the 13 letters mentioned above. Each letter was presented to S at the same optical slant as the typewriter keyboard. The following instructions were given,

I am going to show you some letters, and I would like you to tell me their names. What do you call this one? And this one?

Each response was recorded as it was made.

Training. All Ss were first asked to type a few letters so that they could get used to the task. For those Ss who did not put adequate pressure on the keys, E had them crook the typing finger and practice until their finger pressure was adequate.

The only difference between the Experimental subjects and the Control subjects was the the former typed directly onto a piece of typing paper, and thus could see their own typing product, whereas the latter typed onto a piece of carbon paper backed by a piece of typing paper and therefore could not see their typing product. There was a slight additional difference in that the ribbon lever was not functional for the Control Ss.

The following instructions were given to both groups:

I am going to show you some letters, and I am going to teach you their names. Each time I show you a letter, I will tell you its name, and then I want you to type it. Listen carefully as later on I am going to find out whether you've learned the name.

The first card was then presented, with E pronouncing the name of the letter out loud; at this point further instructions were given,

Now you've got to find the letter. The best thing is to look first along here at each of these letters [E pointed to the top row of letters on the keyboard], and if it is not there look at the next row, then if you don't find it, look along the bottom row.

As soon as S had found and typed the first letter, E presented the second letter and so on, until all 13 letters had been presented. This constituted trial one, and the procedure was repeated until ten trials had been finished. Each trial was timed from the presentation of the first letter to the typing of the 13th letter. At the end of each trial E moved the paper up in the typewriter, and shuffled the cards to ensure a new random order of presentation for the next trial.

Thus all Ss were given the same kind and amount of motor practice, and all had the same kind and amount of perceptual practice (in matching the letter-forms) except for the seeing of self-produced letter-forms by the experimental group.

Post-test. Each child repeated the pretest, and the responses were scored. The Experimental and Control Groups were later compared.

RESULTS

Pretest. The means and variances of the Experimental and Control groups for the pretest are shown in Table 1.

Table 1. Scores for Letter Identification

	Group				
	Experimental (N=10)	Control (N=10)			
<u>Pretest</u>					
Mean Number Identified	4.5	5.5	t= .43	18 df	ns
	S ² =19.2	S ² =23.8	F=1.24	9/9 df	ns
<u>Post-test</u>					
Letter Identification Difference Score	16.0	2.0	t=2.21	18 df	Significant <.05 level

The difference between the means ($t=.43$, 18 df), and the variances ($F=1.24$, 9/9 df) are not significant. This result means that of the initial task the two groups were approximately equal.

Post-test. The post-test was a replication of the pretest, i.e., the task was to name the letters presented by E. A difference score was computed (correct responses on pretest minus correct responses on post-test). The difference score for the Experimental Group was 16 and that for the Control Group was 2 (see Table 1). This difference is significant at the 5 percent level ($t=2.21$, df 18).

DISCUSSION AND SUMMARY

There is some evidence that the Experimental Group learned more than the Control Group since they named a greater number of letters correctly on the post-test than the Control Group did. A comparison of the number of errors made during practice by the two groups shows the same trend (see Table 2 in appendix).

Notwithstanding the small number of Ss used the results suggest that the additional practice in letter-making over and above the association of pretested letters with their names and the association of pretested letters with motor acts as such enabled the experimental subjects to perform better than the controls. The hypothesis was that practice in letter-making ("writing") facilitates letter recognition ("reading"). The Experimental Ss could see their own typing production; the Control Ss were unable to see their own typing production. The results support the hypothesis.

However, Figure 1 shows a slope characteristic of a learning curve for Group B but none for Group C.

(3) The errors were classified for both groups into A) Errors of Similarity (Confusion Errors), and B) Errors of Fecundity (Motor Errors). Rules were formulated for this classification. The analysis showed that of the 189 errors made by both groups, the majority, i.e., 149, were errors of similarity. The Experimental and Control Groups made an almost equal number of errors of proximity (21 and 19 respectively), but the Experimental Group made many fewer errors of similarity (60) than the Control Group (89).

Time Scores

(1) The Experimental Group's time score for the ten training trials was 192 minutes and 30 seconds, compared with the Control

APPENDIX

FURTHER DATA ON LEARNING AND ERRORS

Training

Error Scores

(1) Total Error Scores: The experimental Group made 71 errors and the Control Group made 109 errors for trials 2 to 10; however, this difference is not significant ($t = .95$, $df = 18$).

(2) Error Difference Scores: Error scores were further analyzed by taking the difference scores (as a measure of improvement) between Trial 1 and Trial 10 for both groups. The Experimental Group decreased its total error score by 7 errors and the Control Group had the same number of errors for these trials, i.e., they did not decrease their error score. The difference between the two groups was not significant ($t = 1.02$, $18 df$). (See Table 2.) However, Figure 1 shows a slope characteristic of a learning curve for Group E but none for Group C.

(3) The errors were classified for both groups into A) Errors of Similarity (Confusion Errors), and B) Errors of Proximity (Motor Errors). Rules were formulated for this classification. The analysis showed that of the 189 errors made by both groups, the majority, i.e., 149, were errors of similarity. The Experimental and Control Groups made an almost equal number of errors of proximity (21 and 19 respectively), but the Experimental Group made many fewer errors of similarity (60) than the Control Group (89).

Time Scores

(1) The Experimental Group's time score for the ten training trials was 199 minutes and 50 seconds, compared with the Control

Group's total time of 157 minutes 29 seconds. This difference was significant at the 5 percent level of confidence ($t = 2.04$, $df = 19$).

(2) Time Difference Scores: Time scores were further analyzed by taking the difference score between Trial 1 and Trial 10 as a measure of improvement. The Experimental Group decreased their time scores by 20 minutes and 25 seconds, compared with the decrease in the Control Group's time scores of 18 minutes and 40 seconds. This difference was not significant ($t = .74$, 18 df).

DISCUSSION

A comparison of the Total Error Scores of the Experimental and Control Groups and a comparison of the Error Difference Scores of both groups did not show significant differences; this is probably a result of the small number of errors made. However, the Total Error curves for the two groups (Figure 1) show that there is a difference in the rate of decrease in errors between the groups, with the Experimental Group's curve looking much more like a learning curve than the Control's. The shape of the Experimental Group's curve may be a result of the Experimental Ss learning more about the critical features of letters as a function of the visual feedback they obtained than was possible for the control subjects who did not receive visual feedback. The main errors of similarity were those where M and W, M and N, W and N, and O and D were confused. These confusions are the same as those obtained in much more extensive series of experiments concerned with the critical or distinctive features of letters.

The time curves for both groups show a sharp drop in the course of the training trials, as can be seen from Figure 2. There is, .

however, a persistent small difference between the two groups with the Experimental Group taking longer than the Control Group. The difference is significant at the 5 percent level. This difference in time scores is probably attributable to the fact that the experimental subjects spent time looking at their typed product, whereas the control subjects simply went from typing one letter to looking at the next letter. Time difference scores were computed for both groups, but the Experimental Group showed the larger decrease in time-score. Of all the indices used to measure S's performance, only one did not favor the Experimental Group, suggesting that learning had taken place in this group.

Table 2. Training Scores

	Groups				
	Experimental (N=10)	Control (N=10)			
Total Errors Score	71	109	t= .95	18 df	ns
Error Difference Score	7	0	t=1.02	18 df	ns
Total Time Score (in minutes & seconds)	199.50	157.29	t=2.04	18 df	significant <.05 level
Time Difference Score	29.25	18.40	t= .74	18 df	ns

Figure 1. Errors Per Trial as a Function of Practice

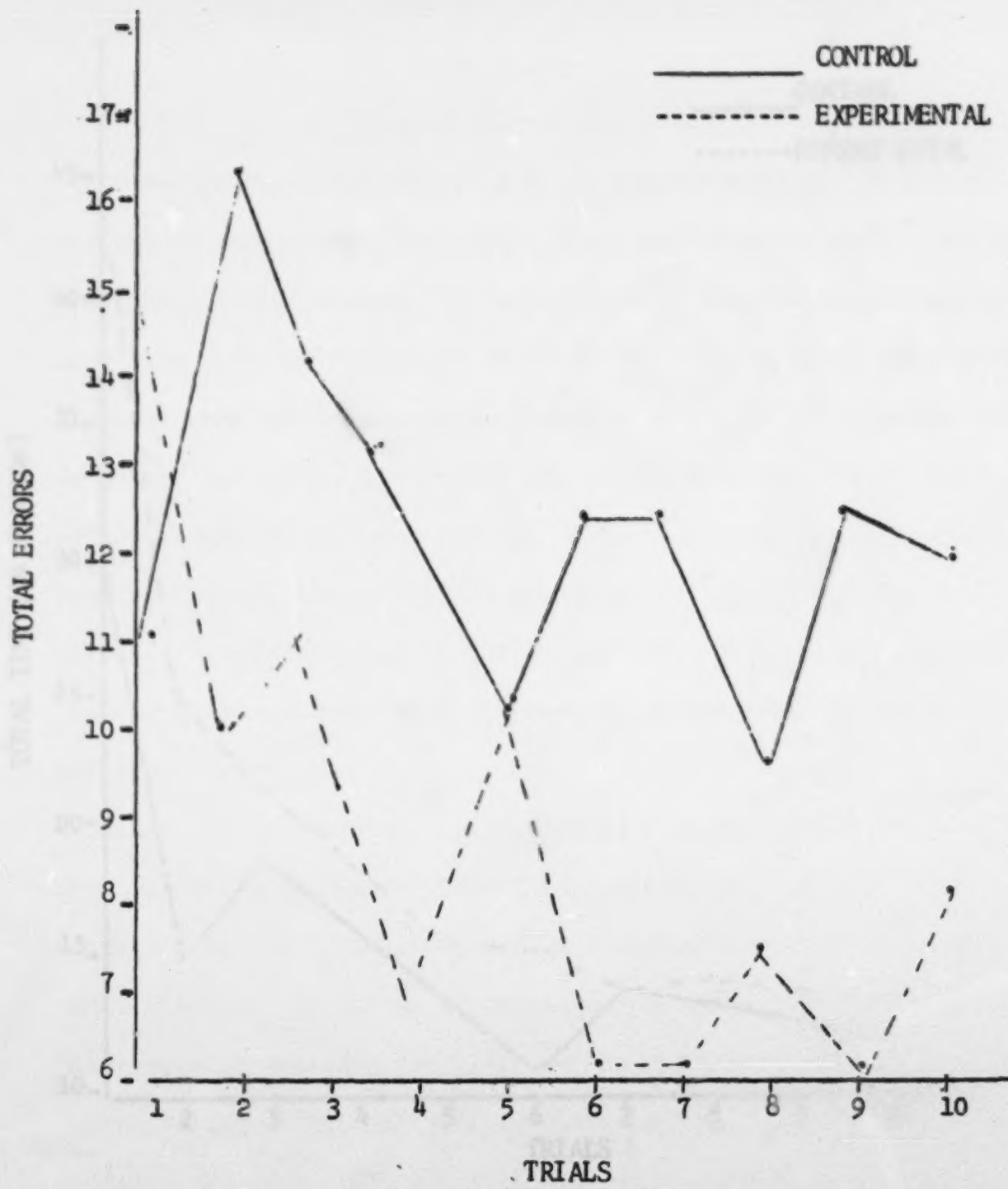


Figure 1. Errors Per Trial as a Function of Practice

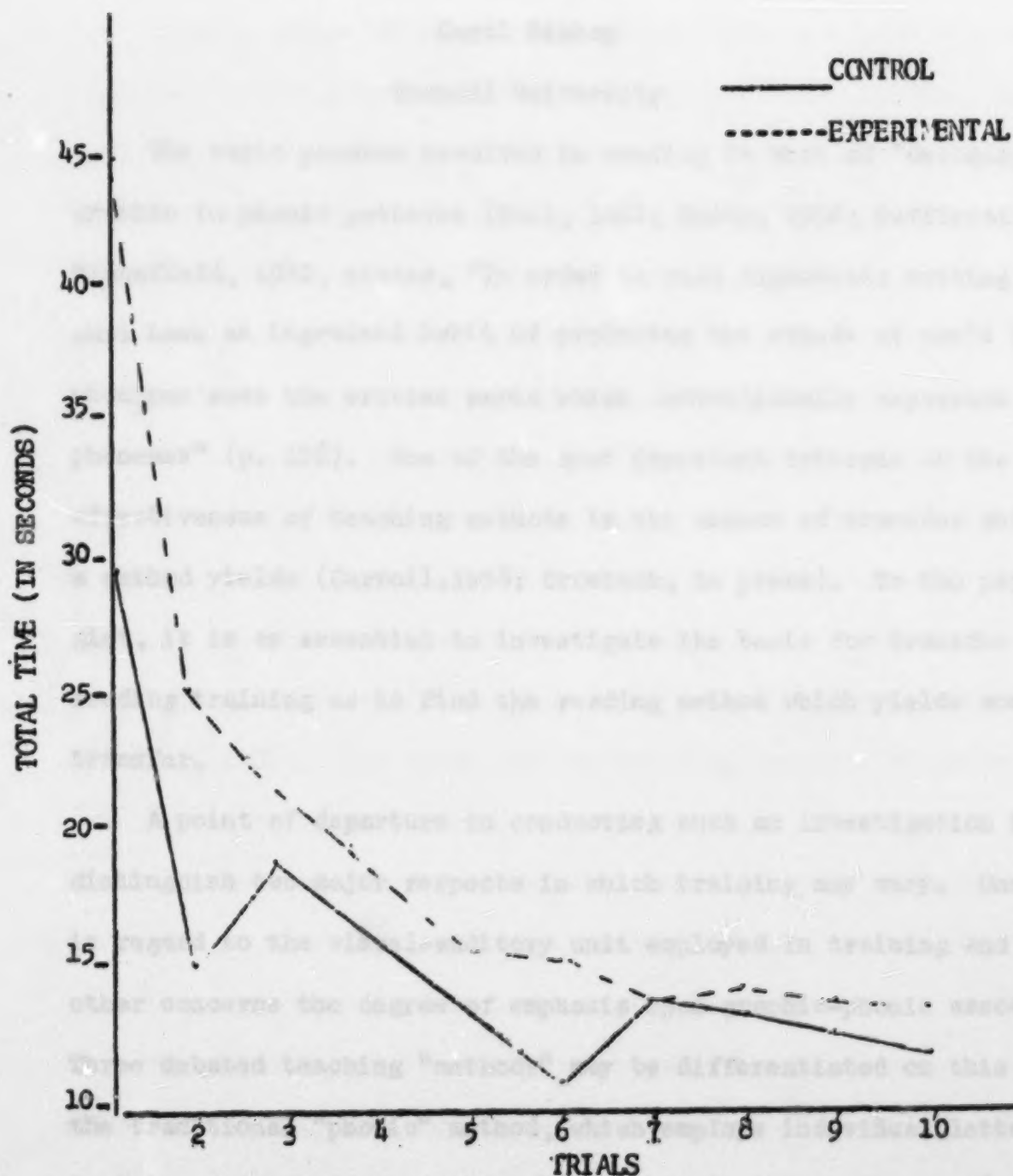


Figure 2. Time Per Trial As A Function of Practice

1. This study is based upon a MA thesis submitted to Cornell University under the direction of Dr. E. J. Gibson. Dr. E. J. Gibson provided technical assistance concerning the Arabic language.

TRANSFER OF WORD AND LETTER TRAINING IN READING¹

Carol Bishop

Cornell University

The basic process involved in reading is that of "decoding" from graphic to phonic patterns (Hall, 1961; Smith, 1956; Soffietti, 1955). Bloomfield, 1942, states, "In order to read alphabetic writing one must have an ingrained habit of producing the sounds of one's language when one sees the written marks which conventionally represent the phonemes" (p. 128). One of the most important criteria of the effectiveness of teaching methods is the amount of transfer which a method yields (Carroll, 1958; Cronbach, in press). To the psychologist, it is as essential to investigate the basis for transfer of reading training as to find the reading method which yields most transfer.

A point of departure in conducting such an investigation is to distinguish two major respects in which training may vary. One is in regard to the visual-auditory unit employed in training and the other concerns the degree of emphasis upon graphic-phonetic associations. Three debated teaching "methods" may be differentiated on this basis: the traditional "phonic" method, which employs individual letters in training and stresses graphic-phonetic relationships, i.e., rules of pronunciation (Dougherty, 1923); the traditional "look-and-say" method, which employs entire words and disregards rules of pronunciation (Huey, 1908); and the more recent Bloomfield system, which stresses

1. This study is based upon a MA thesis submitted to Cornell University under the direction of Dr. E. J. Gibson. Dr. H. M. Hanna provided technical assistance concerning the Arabic language.

graphic-phonetic relationships but recommends training in the form of entire words (Bloomfield & Barnhart, 1961).

Investigations of the transfer effectiveness of the unit employed in training have produced diverse results (Witty & Sizemore, 1955; Gray, 1955). The different findings may be due to the varying relationships among the studies between training unit and degree of emphasis upon graphic-phonetic associations.

It has generally been agreed that a knowledge of graphic-phonetic relationships among units smaller than whole words is necessary for transfer to reading new words. At present, a large-scale research project is being undertaken in England on the assumption that one of the major causes of reading disabilities is the irregularities between spelling and pronunciation in the English language, and the presumably resultant difficulty of acquiring graphic-phonetic relationships (Downing, 1962). Yet there are few empirical answers to questions concerning the role of component graphic-phonetic relationships in reading. Are component graphic-phonetic associations necessary to learn to read? Do they facilitate the process of learning to read, and most important, are they in fact necessary for transfer to new words?

The following experiment was designed to determine the transfer value of training with individual letters as opposed to whole words, and to investigate the role of grapheme-phoneme associations² in reading. The child's experience in learning to read was simulated by teaching adult subjects to read several Arabic words.³ A three-stage transfer design was employed.

-
2. "Grapheme-phoneme associations" will be used to refer to component graphic-phonetic relations, in this case letter-sound relations.
 3. The English translations of the Arabic words were not taught since the associations of meaning to spoken words is a separate learning problem which developmentally precedes reading.

METHOD

Material.

The graphemes used in the experiment were twelve Arabic characters, each with a one-to-one grapheme-phoneme correspondence. The set of letters contained eight consonants and four vowels. These letters were combined to form two sets of eight Arabic words. The letter set and one of the word sets were used for training in Stage II. The other set of words served as the transfer words in Stage III.

The word lists were so constructed that each of the twelve letters appeared at least once in both lists, all of the words were of a CVCV pattern, and half of the consonants appeared twice each as the beginning letters in one list with the other half appearing twice each as beginning letters in the second list (thus preventing word differentiation on the basis of first letters only).

A native speaker of the language recorded on tape the twelve letter-sounds in three different orders and the two sets of words in four different orders. The lists were recorded with a four-second interval between each letter-sound or word, and an interval of twenty seconds between each order.

The graphic form of each letter or word was hand-printed on a 4 x 6 inch index card. The words deviated from true Arabic writing in several respects: (1) the same character represented a sound whether it appeared at the beginning, the middle or the end of a word, (2) the vowel-sounds were always represented by their printed symbols, (3) the letters composing a word were disjoined, and (4) the words were written from left to right. These alterations were introduced to prevent interference from pre-existing reading habits.

The two lists of Arabic words with their phonetic symbols are shown in Figure 1.

FIGURE I .

Arabic Words and Phonetic Symbols

Training Words	Phonetic Symbols*	Transfer Words	Phonetic Symbols
وراف	fa:ru:	یری م	mi:ri:
ی داف	fa:di:	موتوم	mu:tu:
افی ت	ti:fa:	اش وک	ko:ša:
ین وت	tu:ni:	یف اک	ka:fi:
اشی ش	ši:ša:	افی ن	ni:fa:
یمی ش	ši:mi:	یدان	na:di:
یف اد	da:fi:	اشی ر	ri:ša:
وک ود	do:ko:	یفی ر	ri:fi:

*See Pike (1947) for explanation of phonetic symbols.

Subjects.

Sixty students enrolled in a beginning psychology course at Cornell University participated in the experiment to fulfill partial course requirements. The first 40 subjects were alternately assigned to single letter training (Group L) or whole word training (Group W)

when they came for testing. The last 20 subjects received no special training, serving as a control group (Group C). The sexes were equated in all groups.

Procedure.

The procedure was identical for all groups in Stages I and III, but the training procedure in Stage II differed.

Stage I - Phonic Familiarization. All Ss were told that the object of the experiment was to learn to read several Arabic words. Each S became familiar with the transfer set of words by listening to the words played on the tape recorder and by repeating each one aloud in the four-second interval which followed it. Correction was given by E if a word was not pronounced properly. The list was repeated three times.

Stage II - Graphic-Phonic Training.

Group L. S listened to and repeated the twelve letter-sounds. The list was repeated once only, and again, E corrected any errors. S next learned to associate the graphic forms of the letters with their correct sounds. The cards, each with one printed letter, were placed one at a time on a stand in front of S. Each card was exposed for four seconds, at the end of which the correct sound was repeated on the tape recorder.

S was instructed to try to learn the grapheme-phoneme associations on the first repetition of the list, and to respond with the correct sound as soon as he was able, on all subsequent repetitions. The correct sound was always repeated at the end of the four-second response interval.

The list was repeated 16 times in three different orders with an interval of twenty seconds between trials.

Group W. The procedure was identical to that of Group L except that the eight training words were employed instead of the twelve letters. The number of units used in training differed for the groups, and it was decided that the total amount of time spent in training, rather than number of repetitions, was the important factor to equate in order to evaluate the efficiency of the training methods. The word list, therefore, was repeated twice in familiarizing S with the word sounds and was repeated 24 times in training S to read the words, thereby equalling the time spent in training Group L.

Group C. S participated in a second, unrelated experiment which determined recognition thresholds of three-letter English words or nonsense syllables presented tachistoscopically. The amount of time spent in the experiment varied slightly for Ss, but it was approximately the same as that in Groups L and W.

Stage III - Testing for Transfer of Training. All Ss were tested on ability to learn to read the eight transfer words: The cards with the printed words were presented one at a time, and S was to respond with the correct word during a four-second interval. The procedure was identical to that for Groups L and W in Stage II, except that S was instructed to read the words on the first trial if he were able to do so, and the learning trials continued until an individual reached a criterion of two perfect trials.⁴

4. The criterion used in the final analyses was one perfect trial. Each subject was run to two perfect trials in the event that later playback of the tape recording of S's responses would reveal an undetected error on what had been taken as the first perfect trial. This procedure allowed greater assurance of obtaining at least one perfect trial for each subject.

Testing for Grapheme-Phoneme Associations.

At the close of Stage III, all Ss were tested on their ability to give the correct letter-sound within a four-second interval following the presentation of each printed letter. The letters were presented once only.

Subdivision of Training Groups.

The Ss in the two training groups (Groups L and W) were asked to explain how they tried to learn to read the transfer words. Their explanations were tape recorded, and later two judges listened to their answers and decided on the basis of the subjects' reports whether each S learned the words by applying knowledge of individual grapheme-phoneme correspondences ("sounding out" words) or by associating the spoken words with the printed words by some other means. This was a double-blind method in that the Ss did not know the purpose of the experiment when they reported their method of learning the words, and the judges did not know which type of training was given to a particular S nor how well S performed. Any disagreements between judges were discussed, and if agreement could not be reached, S was not categorized.

RESULTS

Measurement of Transfer.

Three criteria of learning were used in Stage III - number of trials to reach perfection, percentage correct of total responses, and number of correct responses on the first transfer trial. The results of comparing the three groups are shown in Table 1. Student "t" tests were applied, and the significance levels were adjusted to compensate for multiple comparisons (Ryan, 1959a; 1959b).

TABLE 1

Effects of Unit of Training in Transfer Stage

	<u>Group L</u>	<u>Group W</u>	<u>Group C</u>
Mean first correct trial	7.35	15.20	24.95
s	5.68	11.14	12.95
t	<u>L vs. C</u> 5.71*	<u>W vs. C</u> 2.62*	<u>L vs. W</u> 2.87*
% correct of total responses	<u>Group L</u> 70.96	<u>Group W</u> 61.21	<u>Group C</u> 48.32
s	16.11	14.78	7.52
t	<u>L vs. C</u> 5.85*	<u>W vs. C</u> 3.51*	<u>L vs. W</u> 2.25*
Mean number of correct responses on 1st trial	<u>Group L</u> 2.90	<u>Group W</u> 1.75	<u>Group C</u> 0.00
s	2.72	2.35	0.00
t			<u>L vs. W</u> 1.47
Mean number of correct letters	<u>Group L</u> 11.85	<u>Group W</u> 8.05	<u>Group C</u> 6.75
s	0.50	3.37	3.20
Range	10-12	1-12	0-10
t	<u>L vs. C</u> 5.62*	<u>W vs. C</u> 1.29	<u>L vs. W</u> 4.07*

* Significant at .05 level per criterion.

Each of the groups is significantly different from the others when trials to perfection and percentage correct of the total responses are compared. The rank order of the groups is Groups L, W and C. Since there were no correct responses on the first trial in Group C, only Groups L and W were tested with such a criterion. The difference

between the two was in the same direction as above, but it was not statistically significant. The failure to find a significant difference on the first trial may have been due to the fact that the letter-trained subjects were seeing the letters in the framework of whole words for the first time. It often took an initial period of adjustment to the new situation before the subjects started to respond.

The learning curves for the three groups are illustrated in Figure 2. Number of correct words is plotted against the average trial on which this number was attained (see Melton, 1935, for method of plotting). The graph shows that throughout the transfer stage, learning to a given criterion of correct words took place in fewest trials for Group L and next fewest for Group W.

The results show, therefore, that letter training has more transfer value than word training, but that word training does produce some transfer.

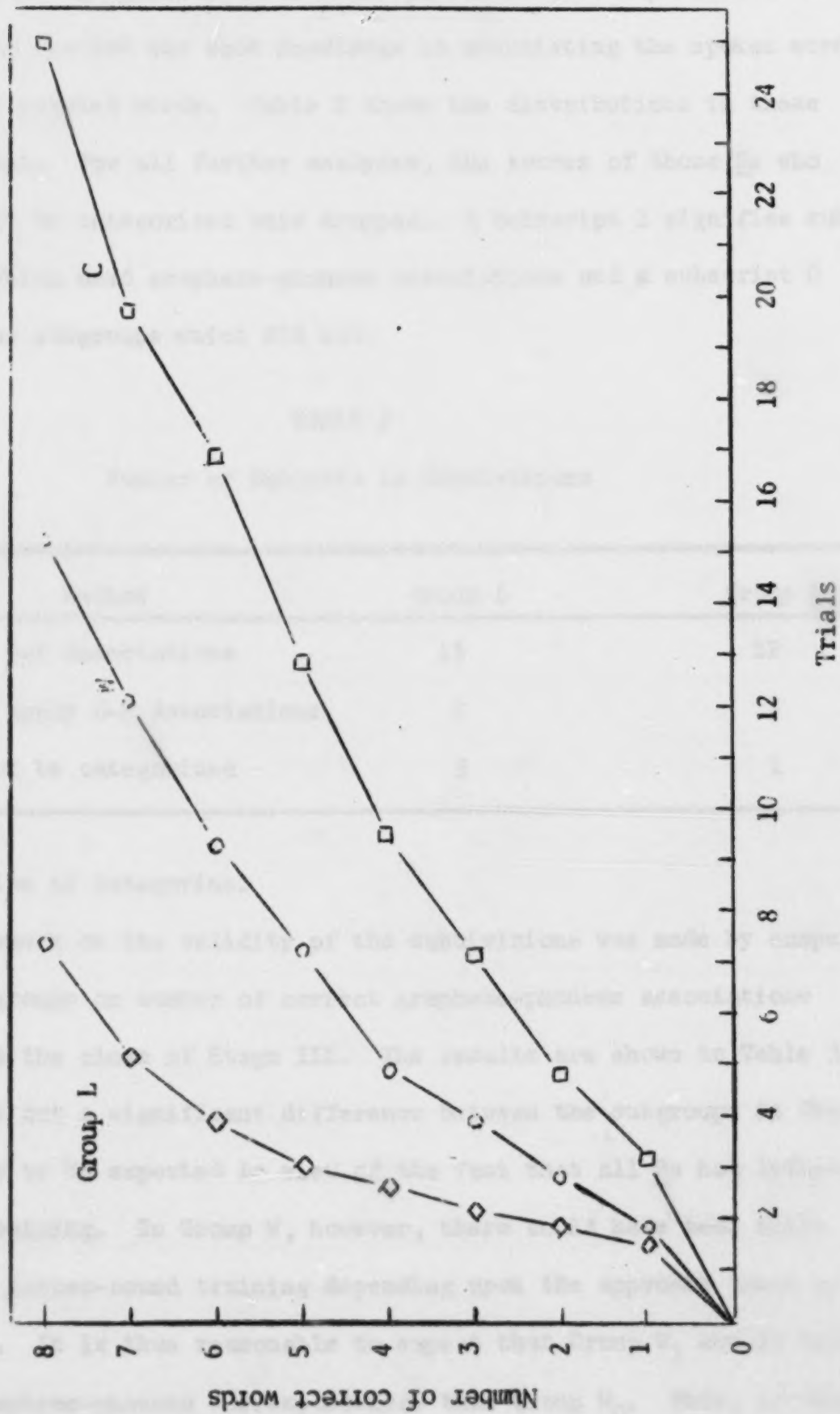
Degree of Learning Grapheme-Phoneme Associations.

The differences between the groups in the average number of letter-sounds known at the end of the transfer period are given in Table 1. Group L knew a significantly greater number of grapheme-phoneme correspondences than either Group W or Group C. There was not a significant difference between the latter two groups; however, some letter-sounds were learned, ranging from 0 - 12 per subject.

Since it was possible to learn grapheme-phoneme associations regardless of the type of training S was assigned, it was necessary to subdivide the training groups in order to evaluate the transfer merits of such associations in reading.

Figure 2

Learning Curves Illustrating the Effects of Training. Number of Correct Words Is Plotted Against Average Trial on Which the Number First was Attained.



Categorization of Subgroups.

Those Ss who learned to read the transfer words by applying knowledge of grapheme-phoneme correspondences were separated from those who did not use such knowledge in associating the spoken words with the printed words. Table 2 shows the distributions in these categories. For all further analyses, the scores of those Ss who could not be categorized were dropped. A subscript 1 signifies subgroups which used grapheme-phoneme associations and a subscript 0 indicates subgroups which did not.

TABLE 2

Number of Subjects in Subdivisions

Method	Group L	Group W
Applied G-P Associations	15	12
Did not apply G-P Associations	2	7
Could not be categorized	3	1

Validation of Categories.

A check on the validity of the subdivisions was made by comparing the subgroups on number of correct grapheme-phoneme associations given at the close of Stage III. The results are shown in Table 3. There is not a significant difference between the subgroups in Group L, which is to be expected in view of the fact that all Ss had letter-sound training. In Group W, however, there could have been self-adopted letter-sound training depending upon the approach taken by a subject. It is thus reasonable to expect that Group W_1 should know more grapheme-phoneme correspondences than Group W_0 . This, in fact,

was true; as is shown in Table 3, a significant difference in the expected direction was found.

TABLE 3

Effects of Grapheme-Phoneme Associations in Transfer Stage

	<u>Group L₁</u>	<u>Group L₀</u>	<u>Group W₁</u>	<u>Group W₀</u>
Mean first correct trial	5.93	20.50	9.75	23.14
s	3.66	5.00	7.15	12.70
t	$\frac{L_1 \text{ vs } L_0}{5.48^*}$		$\frac{W_1 \text{ vs } W_0}{3.16^*}$	
	<u>Group L₁</u>	<u>Group L₀</u>	<u>Group W₁</u>	<u>Group W₀</u>
% correct of total responses	73.00	52.73	68.46	51.03
s	15.43	15.39	14.60	7.33
t	$\frac{L_1 \text{ vs } L_0}{5.81^*}$		$\frac{W_1 \text{ vs } W_0}{3.06^*}$	
	<u>Group L₁</u>	<u>Group L₀</u>	<u>Group W₁</u>	<u>Group W₀</u>
Mean number of correct responses on 1st trial	3.13	0.50	2.83	0.14
s	2.76	1.00	2.51	0.34
t	$\frac{L_1 \text{ vs } L_0}{1.35}$		$\frac{W_1 \text{ vs } W_0}{2.89^*}$	
	<u>Group L₁</u>	<u>Group L₀</u>	<u>Group W₁</u>	<u>Group W₀</u>
Mean number of correct letters	11.87	11.50	10.00	4.86
s	0.53	1.00	1.66	3.38
t	$\frac{L_1 \text{ vs } L_0}{0.86}$		$\frac{W_1 \text{ vs } W_0}{4.80^*}$	

* Significant at .05 level per criterion.

Measurement of Transfer.

The learning curves for the training subgroups are plotted in Figure 3 along with the learning curve for the control group. The curves for Group L and Group W clearly divide into separate curves depending upon the use of grapheme-phoneme associations. The curve for Group L₀ is constructed from the results of two subjects only, so this finding might not be reliable. The word-trained group, however, is subdivided more evenly, and the trend is substantial.

The learning curves reveal little transfer, in trials to learn, for letter-trained subjects who did not apply grapheme-phoneme associations in learning to read the transfer words, and there is essentially no transfer for word-trained subjects who did not apply these associations. On the other hand, the curve for those word-trained subjects who did not apply such knowledge shows substantial transfer, closely resembling that of the letter-trained subjects who used grapheme-phoneme associations to read the words.

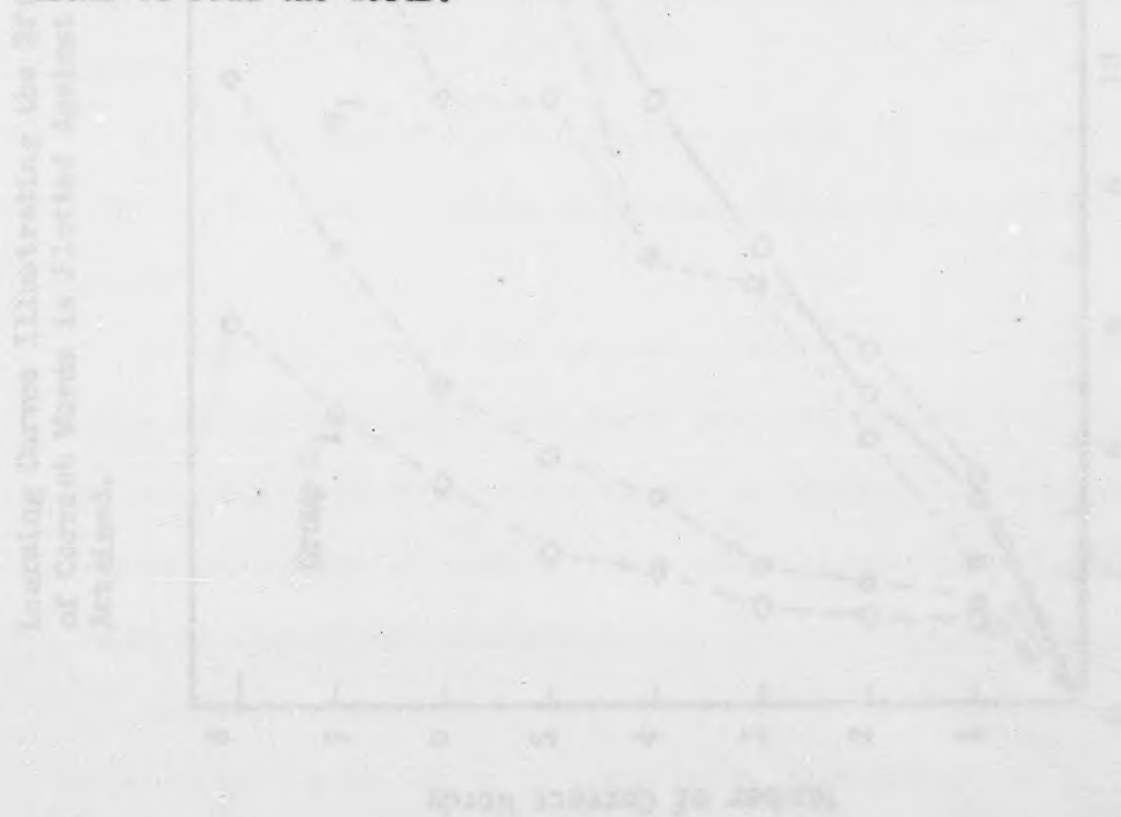
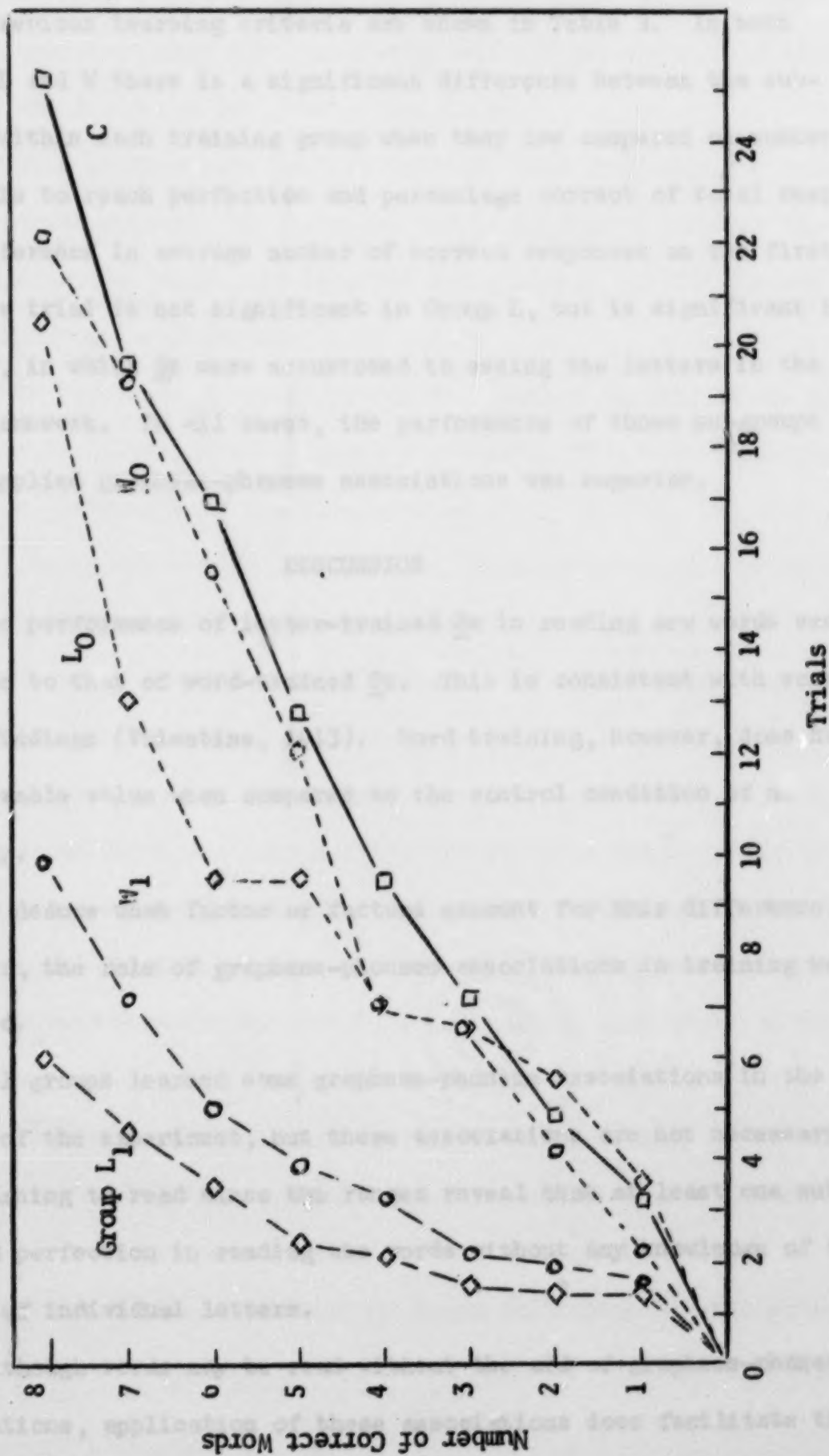


Figure 3

Learning Curves Illustrating the Effects of Grapheme-Phoneme Associations. Number of Correct Words is Plotted Against Average Trial on Which the Number First Was Attained.



The results of comparing the subgroups on the basis of the three previous learning criteria are shown in Table 3. In both Groups L and W there is a significant difference between the subgroups within each training group when they are compared on number of trials to reach perfection and percentage correct of total responses. The difference in average number of correct responses on the first transfer trial is not significant in Group L, but is significant in Group W, in which Ss were accustomed to seeing the letters in the word framework. In all cases, the performance of those subgroups which applied grapheme-phoneme associations was superior.

DISCUSSION

The performance of letter-trained Ss in reading new words was superior to that of word-trained Ss. This is consistent with some other findings (Valentine, 1913). Word training, however, does have considerable value when compared to the control condition of no training.

To deduce what factor or factors account for this difference in transfer, the role of grapheme-phoneme associations in training was analyzed.

All groups learned some grapheme-phoneme associations in the course of the experiment, but these associations are not necessary for learning to read since the ranges reveal that at least one subject reached perfection in reading the words without any knowledge of the sounds of individual letters.

Although words may be read without the aid of grapheme-phoneme associations, application of these associations does facilitate the

learning of new words, i.e., it does have transfer value. With type of training held constant, the performance of subgroups which made use of grapheme-phoneme correspondences was always superior to that of those which did not use such associations.

Grapheme-phoneme associations in fact, were the bases for transfer in this investigation. The learning curves in Figure 3 reveal the very slight effect which training had for subjects who did not use grapheme-phoneme associations.

The two types of training produced different results in the transfer stage, but the differences in group performance were almost entirely due to the difference in percentage of subjects applying grapheme-phoneme associations in the two groups. Letter training provided a propitious opportunity for subjects to form the associations and probably influenced subjects to apply them to reading words. Word training left open the possibility that subjects might set themselves to learn grapheme-phoneme correspondences. It did not have, for individual subjects, transfer value which was only half as effective as letter training (contrary to the apparent results in Figure 2). Instead it had strong transfer value for that portion of the subjects who learned grapheme-phoneme correspondences and little or no transfer effectiveness for the others.

The reason for the overall inferiority of word training as compared to letter training was either that it did not direct as many subjects to learn grapheme-phoneme correspondences or that not all subjects were capable of picking out these relationships when embedded in words. Further research is needed to indicate which alternative is true.

That there are regular (though often complex) graphic-phonetic correlates in the English language, has recently been demonstrated (Hockett, 1960; Venezky, 1962), and that some of these can be learned by children by the end of their first year of reading has also been shown (Gibson, Osser & Pick, in press). Although the present study has dealt with one-to-one grapheme-phoneme correspondences, it is possible that the results of this study might apply to the more global correspondences found by Hockett and Venezky.

This finding would indicate that to obtain a method of teaching children to read which will transfer to new words, further research should be concerned with the most efficient way of teaching children the spelling-to-sound correlates found in the English language. It would be advantageous to test experimentally the transfer values of various methods of teaching these correspondences, and to determine the generality of this transfer for all children of a given age.

SUMMARY

The child's process of learning to read was simulated by teaching adults to read several Arabic words. The purposes of the study were to determine the transfer value of training with individual letters as opposed to whole words, and to investigate the role of grapheme-phoneme associations in reading. It was found that letter training had greater transfer value than word training. Knowledge of grapheme-phoneme correspondences was not necessary to read words, but was necessary for transfer to new words regardless of type of training. The overall inferiority of word training was either that it did not direct as many subjects to learn grapheme-phoneme correspondences, or that not all subjects were capable of picking out these relationships when embedded in words.

REFERENCES

- Bloomfield, L. Linguistics and reading. Elem. English Rev., 1942, 19, 125-130, 183-186.
- Bloomfield, L. & Barnhart, C. L. Let's read, a linguistic approach. Detroit: Wayne State University Press, 1961.
- Carroll, J. B., Chr. Psychological contributions to problems of communication in education, Report of Panel II. Psychological research in education. Washington D.C.: National Academy of Sciences - National Research Council, 1958, 11-18.
- Cronbach, L. J. The psychological background for curriculum experimentation. Rosenbloom, Paul C., Ed., McGraw-Hill, in press.
- Dougherty, Mary L. How to teach phonics. New York: Houghton-Mifflin Co., 1923.
- Downing, J. A. The relationship between reading attainment and the inconsistency of English spelling at the infants' school stage. Brit. J. educ. Psychol., 1962, 32, 166-178.
- Gilson, Eleanor J., Osser, H. & Pick, Ann D. A study of the development of grapheme-phoneme correspondences. J. verb. learn. verb. Behav., in press.
- Gray, W. S. Phonic versus other methods of teaching reading. Read. Teacher, 1955, 9, 102-106.
- Hall, R. A. Jr. Sound and spelling in English. Philadelphia: Chilton Co., 1961.
- Hockett, C. N. Analysis of English spelling, Part I: Analysis of graphic monosyllables. Unpublished manuscript, 1960.
- Huey, E. B. The psychology and pedagogy of reading. New York: Macmillan, 1908.
- Melton, A. W. The end-spurt in memorization curves as an artifact of the averaging of individual curves. Psychol. Monogr., 1935, No. 212.
- Pike, K. L. Phonemics: A technique for reducing languages to writing. Vol. 3. Linguistics. Ann Arbor: University of Michigan Publications, 1947.
- Ryan, T. A. Multiple comparisons in psychological research. Psychol. Bull., 1959a, 56, 26-47.
- Ryan, T. A. Erratum. Psychol. Bull., 1959b, 56, 396.

- Smith, H. L., Jr. Linguistic science and the teaching of English.
Cambridge: Harvard University Press, 1956.
- Soffietti, J. P. Why children fail to read: a linguistic analysis.
Harv. educ. Rev., 1955, 25, 63-84.
- Valentine, C. W. Experiments on the method of teaching reading.
J. exper. Ped., 1913, 2, 99-112.
- Venezky, R. A computer program for deriving spelling-to-sound
correlations. Unpublished master's thesis, Cornell University,
1962.
- Witty, P. & Sizemore, R. Phonics in the reading program: a review and
an evaluation. Elem. English Rev., 1955, 32, 355-371.

THE ROLE OF GRAPHEME-PHONEME CORRESPONDENCE IN
THE PERCEPTION OF WORDS*

By Eleanor J. Gibson, Anne Pick, Harry Osser,
and Marcia Hammond, Cornell University

What is the critical unit of language for the reading process-- the perception of written words? Recommendations for the teaching of reading have usually depended on the choice and definition of a unit. It was once assumed that the letters of the alphabet were the units, so children began reading instruction by learning their ABCs. The advent of Gestalt psychology (applied in a very simple-minded way) changed all this. Several generations of children, as a result, were taught by the 'whole word' method. As Bloomfield points out, such a method loses all the advantages of an alphabetical language.¹ One might as well learn picture-writing. Yet the letters of our alphabet in many cases do not have a one-to-one correlation with phonemic units. Is there, then, a proper unit?

It is the purpose of this paper to investigate the hypothesis that a unit is constituted by spelling-to-sound correlations. This is a higher-order unit formed by grapheme-phoneme correspondences. It is not defined by speech alone, or writing alone, nor is it a morpheme. It is the letter-group which has an invariant relationship with a phonemic pattern. Whole words usually have such a relationship; but often they can be broken into smaller clusters of letters

*Published in the Amer. J. Psychol., 1962, 75, 554-570.

1. Leonard Bloomfield, Linguistics and reading, Elementary Engl. Rev., 19, 1942, 125-130, 183-186.

which still have the kind of relationship referred to when they are in a stated position relative to other such clusters. The clusters may be of different sizes and the rules for the grapheme-phoneme correspondence are conditional on what precedes or what follows. The correspondence is not a matter of 'sequential probability' but it is to be found in the structure of written English as it is related to spoken English.

The existence of such correspondence has been demonstrated in a study by Charles F. Hockett and a group of co-workers in the Department of Linguistics at Cornell University. This group has classified English graphic monosyllables (1) according to the arrangements of letters in them, and (2) in terms of their pronunciation. The aim of the classification is to discover the rules by which pronunciation can be predicted from spelling. A graphic monosyllable includes a vowel-spelling, an initial-consonant spelling, and a final-consonant spelling. An initial-consonant spelling might be CL; a vowel-spelling, EA; a final consonant spelling, TS. The monosyllable is CLEATS. Rules for pronunciation are formulated conditionally on what precedes and follows as well as in terms of the letter-group as such. Regularities in spelling-to-sound predictability may be termed spelling-to-sound correlations, or grapheme-phoneme correspondences.

A child who learns to read (well) is forming useful spelling-to-sound habits based on these rules, whether he could tell you so or not. Even if he is taught by some method which makes it difficult, he must eventually discover the important spelling-to-sound correlations, if he is to be able to generate for himself the way to read new words.

From this starting position, we can proceed to the statement of a psychological hypothesis which has some testable consequences. Reading consists of decoding graphic material to the phonemic patterns of spoken language which have already been mastered when reading is begun. The units to be decoded are not single letters, for these have no invariant acoustic match in our language. The whole word is possible, but is uneconomical as a training unit for it provides no basis for independent decoding of new graphic combinations. The hypothesis advanced is that the reading task is essentially that of discovering higher-order invariants, the spelling-to-sound correlations. These are constants which are presumably discovered by exposure to both the graphic and the phonemic stimuli at the same time and in different contexts, so that the invariant combinations can be recognized in many different words. There are not a fixed number of units, such as all the words in the dictionary. There are, rather, all the potential ones which the rules of correspondence generate.

It is assumed that the individual discovers these grapheme-phoneme correspondences as he learns to read, even if he is not specifically taught them. He may not be able to draw up a set of rules but he 'has' them, if he is a good reader, just as a young child 'has' grammar long before he can formulate rules for it. Once he has them, any letter-combinations which follow the rules are functional units.

From this hypothesis we predict that a skilled reader should discriminate better visually, letter-groups (pseudo-words) which are constructed according to the rules of spelling-to-sound correlation

(found in the structure of written and spoken English) than ones which are not, or are only partially so. This hypothesis is tested in the experiments described below. Recognition of letter-groups presented tachistoscopically was the criterion chosen for discrimination. Ability to perceive a letter-group presented tachistoscopically was found by Crausman to correlate with reading ability.² There is reason to think, therefore, that it is a valid measure of word-discriminability in the reading process.

HISTORICAL BACKGROUND

There is a certain amount of relevant material in the literature of experimental psychology which it may repay us to examine in the light of this hypothesis. Particularly, a number of experiments on word- and letter-recognition have been conducted in recent years with the method of tachistoscopic viewing. Howes and Solomon said, in reporting one of these, "any general property of words . . . should be operative in the word-threshold experiment, and the dependent variable (duration threshold) some function of it."³ What is the logic of this statement, and what are the properties of words (or letter groups) which have been treated as parameters in a 'word-threshold' experiment?

The logic is discussed by Solomon and Howes. Perception, they say, is defined by both stimulus-properties and response-properties. When linguistic responses are elicited in a perceptual experiment, "any variable that is a general property of linguistic responses must

2. Burt Crausman, Temporal perception of good and poor readers, Canad. Psychol., 1, 1960, 113 (abstract).

3. D. H. Howes and R. L. Solomon, Visual duration threshold as a function of word probability, J. exp. Psychol., 41, 1951, 401.

also be a property of any perceptual concept that is based upon these responses."⁴ Word-frequency was the property investigated by Solomon and Howes. They defined it simply in terms of frequency in the Thorndike-Lorge tables, but pointed out that frequency could refer to "frequency of visual exposure of a linguistic stimulus" or "the frequency of occurrence of a linguistic response." The two are different, they thought, but probably correlated. The assumption has been expressed by Postman and Conger that emitted frequency of a word, as such, affects the visual cognitive threshold.⁵ Osgood, on the other hand, questions this assumption: "frequency-of-conversational-usage would not affect recognizability of words as visual stimuli."⁶

Postman and Rosenzweig found, in line with Osgood's position, that auditory training (hearing and speaking) did not significantly decrease visual cognitive thresholds for printed syllables, although it did decrease auditory thresholds, and visual training decreased visual thresholds.⁷ Sprague also found that oral practice alone had no effect on tachistoscopic cognitive thresholds.⁸ On the other hand, Forrest found that visual thresholds (determined by increasing

4. R. L. Solomon and D. H. Howes, Word frequency, personal values, and visual duration thresholds, Psychol. Rev., 58, 1951, 257.

5. Leo Postman and Beverly Conger, Verbal habits and the visual recognition of words, Science, 19, 1954, 671-673.

6. C. E. Osgood, Method and Theory in Experimental Psychology, 1953, 293.

7. Leo Postman and M. R. Rosenzweig, Practice and transfer in the visual and auditory recognition of verbal stimuli, this JOURNAL, 69, 1956, 209-226.

8. R. L. Sprague, Effects of differential training on tachistoscopic recognition thresholds, J. exp. Psychol., 58, 1959, 227-231.

sharpness of focus) varied with relative familiarity of paralog, although the method of familiarizing was by auditory presentation.⁹ (It should be noted that the paralog were all "pronounceable" and that they were exhibited twice in written form so that S would be familiar with the spelling of the words.)

The hypothesis proposed in this paper attempts to cut across the dichotomy of stimulus- and response-frequency by proposing that the critical unit to be considered in word-recognition involves both--that it is, in fact, a stimulus-response correspondence which the individual becomes skilled in detecting as it recurs in an invariant relationship in different words. Neither frequency of visual exposure alone or emitting the oral response alone is of significance. Rather, it is frequency of experiencing a grapheme-phoneme coincidence which leads to skilled recognition.

No experiments have been performed heretofore with this concept in mind, but the experiments which have attempted to relate some property of words or letter-groups to visual discriminability will be briefly summarized.

Word-frequency. The hypothesis that greater frequency of words in the language should be correlated with speed and accuracy of recognition has been verified in a number of experiments. Cattell's early experiments with groups of letters and words exposed on a revolving drum laid the foundation for this research,¹⁰ followed by Erdmann and

9. D. W. Forrest, Auditory familiarity as a determinant of visual threshold, this JOURNAL, 70, 1957, 634-636.

10. J. McK. Cattell, Ueber der Zeit der Erkennung und Benennung von Schriftzeichen, Bildern, und Farben, Philos. Stud., 2, 1885, 635-650.

Dodge.¹¹ Vernon,¹² Howes and Solomon,¹³ Solomon and Postman,¹⁴ McGinnies, Comer and Lacey,¹⁵ Postman and Conger,¹⁶ King-Ellison and Jenkins,¹⁷ Baker and Feldman,¹⁸ and Moscovici and Humbert¹⁹ have all confirmed the role of word-frequency (in some cases built in experimentally) in lowering visual recognitive thresholds.

Experimental findings for 'familiarity' of words are comparable. Using Noble's technique of scaling words for familiarity by means of a rating scale,²⁰ Kristofferson obtained recognition-thresholds for

11. Benno Erdmann and Raymond Dodge, *Psychologische Untersuchungen uber das Lesen auf experimenteller Grundlage*, 1898, 360 pp.

12. M. D. Vernon, *Studies in the Psychology of Reading: A. The Errors Made in Reading*, Med. Res. Council, Spec. Rep. Ser., No. 130, 1-40.

13. Howes and Solomon, 1951, *op. cit.*, 401-410.

14. R. L. Solomon and Leo Postman, Frequency of usage as a determinant of recognition thresholds for words, *J. exp. Psychol.*, 13, 1952, 195-201.

15. E. M. McGinnies, P. B. Comer, and O. L. Lacey, Visual recognition thresholds as a function of word length and word frequency, *ibid.*, 44, 1952, 65-69.

16. Postman and Conger, *op. cit.*, 671-673.

17. Patricia King-Ellison and J. J. Jenkins, The durational threshold of visual recognition as a function of word frequency, *this JOURNAL*, 67, 1954, 700-703.

18. K. E. Baker and Herman Feldman, Threshold-luminance for recognition in relation to frequency of prior exposure, *this JOURNAL*, 69, 1956, 278-280.

19. S. Moscovici and T. C. Humbert, Usage et disponibilite comme facteurs determinant la duree de stimuli verbaux, *Bull. de Psychol.*, 13, 1960, 406-412.

20. C. E. Noble, The meaning-familiarity relationship, *Psychol. Rev.*, 60, 1953, 89-98.

20 items, some real and some pseudo-words, and found a significant correlation between rated familiarity and recognition-threshold.²¹

A developmental study by Hoffman compared children in grades 1 through 8 and adults on tachistoscopic viewing of (1) groups of consonants, (2) nonsense-syllables (pronounceable), (3) unfamiliar words, and (4) familiar words.²² The number of letters read correctly increased from condition (1) through (4). The developmental curves differ, as well. The slope is least steep for consonants and increases with the conditions. In other words, groups of consonants are not recognized much better by adults than by first-graders, but the advantage increases as approximation to familiar linguistic units increases.

Miller, Bruner, and Postman compared accuracy of visual recognition for groups of letters having four orders of 'approximation to English.'²³ Accuracy of recognition was greater as approximation to English increased. That frequency as such of either single letters, bigrams, etc., is the critical principle here is unlikely. When the four lists are examined, it is obvious that pronunciability increases

21. A. B. Kristofferson, Word recognition, meaningfulness, and familiarity, Percep. mot. Skills, 7, 1957, 219-220.

22. Jakob Hoffman, Experimentell-psychologische Untersuchungen uber Leseleistungen von Schulkindern, Arch. gesam. Psychol., 58, 1927, 325-388.

23. G. A. Miller, J. S. Bruner, and Leo Postman, Familiarity of letter sequences and tachistoscopic identification, J. gen Psychol., 50, 1954, 129-139.

with order of approximation to English, and with each stage, more and more 'real' syllables appear. Pronounceable and familiar sub-groups may have played a very important role in increasing accuracy.

Trigram-frequency. Postman and Conger investigated the effect of trigram-frequency (sequences of three letters which appear as parts of words with different frequencies) on recognition-threshold.²⁴ Two experiments failed to show any relationship between recognition-thresholds and trigram-frequency.

An experiment by Taylor investigated the effect of experimentally produced frequency on trigrams.²⁵ The trigrams were chosen so as to be approximately equal for summed (individual-letter) frequency. They were selected also for easy pronunciability. An experimental list was presented to the Ss 15 times. S looked at each letter-group for two seconds and pronounced it aloud. Afterward, duration-thresholds were obtained for these and for a matched control-list. Frequency of practice of this type did lead to a lowered recognition-threshold. Thus, when a letter-group is viewed and pronounced as a unit, frequency is effective.²⁶ But mere frequency of sequences embedded in words is not.

Trigrams which form sub-units (that is, pronounceable syllables) of words are another matter. Postman and Rosenzweig compared

24. Postman and Conger, op. cit., 671-673.

25. J. A. Taylor, Meaning, frequency, and visual duration threshold, J. exp. Psychol., 55, 1958, 329-334.

26. See also Solomon and Postman, op. cit., 195-201.

recognitive thresholds for sets of trigrams which were high and low, respectively, for frequency of occurrence both as actual syllables in English words and as trigrams.²⁷ The more 'familiar' syllables had lower mean recognitive thresholds. In fact, English words yielded no lower thresholds than familiar syllables.

Meaningfulness. Whether 'meaningfulness' has any role in word-recognition has been the subject of a few experiments. The earliest of these by Peixotto did not use a short-exposure technique, but the results are relevant.²⁸ She used syllables scaled for associative-value from the lists of Glaze,²⁹ Hull,³⁰ and Krueger.³¹ The Ss simply read a list of 30 syllables printed in columns on pages of a booklet. After reading through three pages (the same list repeated in random order) they found syllables they had read mixed with 90 others on the fourth page. They marked the ones they recognized as having seen before. She found little or no relationship between associative value and recognitive score.

27. Postman and Rosenzweig, op. cit., 209-226.

28. H. E. Peixotto, The recognitive value of three hundred nonsense syllables, this JOURNAL, 61, 1948, 352-360.

29. J. A. Glaze, The associative value of nonsense syllables, J. genet. Psychol., 35, 1928, 255-267.

30. C. L. Hull, The meaningfulness of 320 selected nonsense syllables, this JOURNAL, 45, 1933, 730-734.

31. W. C. F. Kruger, The relative difficulty of nonsense syllables, J. exp. Psychol., 17, 1934, 145-153.

A different criterion of meaning was used by Taylor, who had Ss study syllables accompanied by a picture of a familiar object which was designed as their 'meaning.'³² Recognitive thresholds were not lowered for this group as compared with one which had equal frequency of practice with the syllable without the picture (see above). Neither did the degree of associative value of the syllable (as determined by Glaze) have any effect upon performance. She concluded that the variable of meaningfulness, as defined by these procedures, did not influence visual thresholds of duration.

Kristofferson³³ found a significant correlation between meaning (as scaled by Noble³⁴) and threshold for word-recognition. In this case, however, meaning covaried with 'familiarity.' It is probable that the latter variable was the effective one, in view of the usual positive results with word-frequency and the negative ones for meaningfulness.

The finding of Postman and Rosenzweig that three-letter English words were not recognized any sooner than familiar syllables (sub-units) of words is consistent with the rejection of 'meaning' as a variable in word recognition.³⁵ Their words all had a referent (e.g., bun, pen, etc.), whereas the syllables did not (e.g., ing, est, etc.). They commented that "the failure of the English words to yield lower

32. Taylor, op. cit., 329-334.

33. Kristofferson, op. cit., 219-220.

34. Noble, An analysis of meaning, Psychol. Rev., 59, 1952, 421-430.

35. Postman and Rosenzweig, op. cit., 216.

thresholds than the nonsense-syllables suggests that S is no less ready to use syllables as response-units than he is English words of comparable linguistic frequency."

Summary. These experimental results taken together seem to us consistent with our hypothesis that letter-groups which have an invariant spelling-to-sound relationship form functional units for reading. Frequency in one sense would be a necessary condition for the formation of such units, for invariance cannot be discovered without repetition. It would be expected, therefore, that whole words of high frequency and syllables (sub-units of words having a consistent pronunciation) would have high recognitive values, whereas mere sequential frequency would not unless it had a consistent acoustic match. The evidence to be drawn from these experiments is, however, merely supportive; a real test of the hypothesis remains to be made.

PRELIMINARY EXPERIMENT

The purpose of the experiment was to test the hypothesis that skilled readers would discriminate visually pseudo-words constructed according to the rules of spelling-to-sound correlation better than words which are not, or are only partially, so constructed. Three lists of pseudo-words were constructed, one with an invariant relation between spelling and sound (e.g., NOOSH), one in which all the words were pronounceable, but there were two alternative pronunciations (e.g., DRIEND), and one in which the words had low spelling-to-sound correlation (were relatively unpronounceable by rules of English pronunciation, e.g., SCRIGW). Word-length varied, but was equated for the three lists. There were 15 words in each list.

The words were presented tachistoscopically, one at a time for 100 m.sec. each. The three lists were mixed in a random order. Items were projected on a screen so that Ss could be run in groups. (A second projector and tachistoscope, synchronized with the first, provided a pre- and post-exposure field of a brightness equal to that of the background of the projected word.) The pre-exposure field contained two horizontal lines so arranged as to frame the projected word if they were superposed on it. These lines took the place of a fixation-point. The slides containing the words were prepared by typing in bold-face type with a sign-typewriter (typed letters were 3/8 inches high).

The Ss were given a 'ready' signal, fixated the lines, and after the word was flashed, tried to write what they saw. They were told to guess when they could. Each word was exposed only once. There were 21 Ss, members of a class in the Psychology of Language.

Results were scored right or wrong for the whole word (scoring of parts was tried but discarded as being difficult and no more informative). The mean number correct was about twice as great for the two pronounceable sets of words as for the unpronounceable. A Tukey-test for multiple comparison of means showed that the means for the two pronounceable series were both different from the unpronounceable, but not from each other. If these two had differed, it might have been hypothesized that two alternate pronunciations interfere with one another and reduce discriminability in comparison with a single 'good' pronunciation. This was not the case.

EXPERIMENT I

The hypothesis tested in Experiment I was the same as in the preliminary experiment, but only two sets of words were constructed: one with high spelling-to-sound correlation (referred to as the P or pronounceable list) and one with low spelling-to-sound correlation (the U or unpronounceable list). Summed letter-frequency, which has not been controlled in the preliminary experiment, was equated for the two lists.

Construction of lists. The 25 words in the pronounceable list consisted of (1) an initial consonant-spelling having a single regular pronunciation, (2) a final consonant-spelling having a single regular pronunciation, and (3) a vowel-spelling placed between the two consonant-spellings, and having a single regular pronunciation when it follows and is followed by the given initial and final consonant-spellings, respectively. An example would be B/1, E/3, SKS/2. Some of the consonant- and vowel-spellings were single letters, some were 'clusters.' A cluster-type of vowel-spelling may actually contain a consonant-letter as in 'aw.'³⁶

The words in the second list--those with low spelling-to-sound correlation--were constructed from the words in the first list. The initial consonant-spelling of the pronounceable words became the final consonant-spelling of the unpronounceable words. Similarly, the final consonant-spelling of the pronounceable words became the

³⁶. Spelling must often be 'mapped' into phonemes by clusters or groups of letters, rather than by single letters. Rules for invariance are thus phrased in terms of 'consonant-spelling' and 'vowel-spelling.'

initial consonant-spelling of the unpronounceable words. The position of the vowel spellings remained the same.³⁷ For example: Pronounceable, B E SKS; Unpronounceable, SKS E B. The complete list of words appears in Table I.

TABLE I

Correct Responses and Pronunciability-Ratings for
Pronounceable (P) and Unpronounceable (U) Words

P Words	Total Correct	Pronun.- Rating	U Words	Total Correct	Pronun.- Rating
(1) DINK	118	1.06	NKID	105	5.85
(2) VUNS	110	1.53	NSUV	90	6.34
(3) GLOX	117	1.59	XOGL	108	6.12
(4) SULB	95	1.95	LEUS	115	5.64
(5) LODS	107	1.28	DSOL	118	4.99
(6) BESKS	79	2.30	SKSEB	50	6.77
(7) CLATS	91	1.47	TSACL	102	6.05
(8) BRELP	91	3.01	LPEBR	35	7.72
(9) FRAMB	112	2.40	MBAFR	36	7.31
(10) GRISP	113	1.62	SPIOR	77	5.24
(11) FUNTS	119	1.62	NTSUF	77	6.10
(12) TILMS	97	2.18	LMSIT	70	6.44
(13) SLAND	120	1.71	NDASL	69	6.73
(14) BLASPS	67	2.88	SPSABL	35	7.28
(15) SPRILK	38	3.21	LKISPR	14	7.40
(16) SMAWMP	3	6.20	MPANEM	20	7.23
(17) BLORDS	54	2.29	DSORBL	49	6.00
(18) PREENT	56	2.86	NTEEPB	7	7.47
(19) DRIGHK	36	6.34	KIGHDR	0	6.72
(20) KLERFT	37	4.18	FTERKL	33	6.77
(21) GLURCK	51	2.94	CKURGL	37	6.59
(22) QUEESK	45	3.05	SKEEQU	35	4.63
(23) TIRPTH	42	4.50	PTHIRT	31	6.12
(24) PRILTHS	28	4.63	LTHSIPB	0	8.36
(25) TRILFTHS	13	5.29	LFTHSITB	0	8.50

37. To obtain a pronounceable and an unpronounceable list using the same letters, it was necessary to choose initial consonant-spellings that are not also final ones, and vice versa; or at least one of the consonant-spellings in each pair had to be of this type.

The reader may ask, at this point, whether one list is, in fact, pronounceable while the other is not. Actually, the term pronounceability is used here as a convenient means of referring to a correlation of spelling-with-sound, according to the rules of English spelling and pronunciation. One can make an attempt to pronounce 'TSACL' some way or other, but there is no invariant correlation between this spelling and a standard way of pronouncing it in the English language.

Procedure. The task required of S was to identify a word as it was projected on a screen, and to write it down. The procedure was similar to the preliminary experiment in the method of projection and apparatus employed; but it differed in that five presentations of the lists were given, with an increasing exposure-time for each presentation.³⁸ The 50 words were projected in a random order for 30 m. sec. each. They were then reordered and projected for 50, 100, 150, and 250 m. sec. The words were reordered after each series. The Ss were seated about four feet from the screen and were told to write down what they saw and to guess when they could. They were run in small groups to permit good viewing angles for all Ss. The height of the letters projected on the screen was three and three-fourths inches, the width about two and a half inches. The entire field of projection was 82 inches wide. There were 25 Ss (adult, summer-school students); their native language was English.

38. This technique was copied from Miller, Bruner, and Postman, op. cit., 129-139.

Results. The words recorded by the Ss were scored right or wrong, and separate means for P and U words were obtained for each time-interval. The mean percentage of P words correct is consistently higher than that for U words. The percentage correct rises sharply from the first exposure to the second, and thereafter increases very little. The two curves are parallel throughout.

The Wilcoxon Matched Pairs Signed Ranks Test³⁹ was used to compare the total P and U scores (see Table I). The pairs in each case were composed of the two pseudo-words constructed from the same group of letters, one pronounceable, one not. The difference between the total P score and the total U score is significant at better than the 1 percent level (two-tail).

It was also possible to test the consistency of the trend for Ss, with Wilcoxon's test. When the scores (number correct) of the P and U words were compared for the Ss, the difference between them was significant beyond the 1 percent level. The difference between them at every time-interval was also significant beyond the 1 percent level (two-tail test in all cases).

Because of a long established habit of reading from left to right, it might be expected that more errors would occur for the final consonant-spelling than for the initial one. The Ss' reproductions were scored for initial and final consonant-clusters separately, and this expectation was borne out. But the difference between words of the P and U lists was still present ($p < .001$ by a

39. F. Wilcoxon, Some Rapid Approximate Statistical Procedures, 1949

chi-square test for both initial and final clusters).

Errors can also be expected to increase with length of word. Such was the case. Errors were fewest in the case of four-letter words and increased with length. The difference between P and U words occurred for all lengths of word but was smallest for four-letter words and greatest for five-letter words. The interaction could not be measured, since the number of cases for the different lengths of word was not equal, but the difference is quite obvious. A short word of four letters is so seldom missed that the difference between P and U can have only a small effect. On the other hand, a word of eight letters is so difficult that not even pronunciability will permit perfect discrimination to occur with great frequency.

The errors were studied and classified also for the kinds of changes made. "Real" word-errors occurred infrequently, probably because the Ss were told in the instructions that they would be shown 'nonsense' words. More real words were given for P combinations than for U combinations (27 in all, as against 15).

Errors which changed the projected letter-group in the direction of a more pronounceable one were frequent. This was accomplished in some cases by adding a vowel (e.g. NIKID for NKID), by omitting a consonant (e.g., SKEB for SKSEB), or by changing a consonant-cluster (e.g., BLUS for LEUS). An error of this type was often repeated on later trials.

Errors of omission (writing nothing) were more frequent for the unpronounceable words, but such errors did not occur often. They occurred most often for the longer words.

Discussion. The results of this experiment demonstrate that a letter-group with a high spelling-to-sound correlation is reproduced more accurately than an equivalent letter-group with a low spelling-to-sound correlation. This result cannot be caused by a difference in the familiarity of the letters taken alone, or even the vowel- and consonant-clusters taken alone, for the same clusters were used in the two lists. It must be due to the existence of higher-order graphic units: the letter-combinations of English writing that function as relatively stable units in grapheme-phenome correspondence.

Practically, this result suggests strongly that the proper unit for analyzing the process of reading (and writing) is not the alphabetical letter but the spelling pattern which has an invariant relationship with a phonemic pattern. This may be of great importance for children's learning to read and write.

Among the theoretical implications for psychology, perhaps the most general is this, that while reading is based on the discrimination and identification of visual forms such as letters, it becomes, in the skilled reader, a process of perceiving 'super-forms,' and that these tend to be constituted (organized) by their relation to auditory-vocal temporal patterns. Insofar as frequency has a role in the constitution of these units, it is the frequency of grapheme-phoneme coincidence which is crucial, not frequency of exposure to the seen or uttered units alone. The reading of words is thus inseparable from the hearing of words. Since the hearing of words is also inseparable from the speaking of words, reading must be

conceived, however, as part of a circular response-process, not simply as a stimulus-response process. The stimuli are complex, and the eventually effective stimuli are response-produced, in the sense that the visual stimuli which come to function as units are structured partly by feedback from response. Verbal perception, unlike certain other kinds of perception, will have to be explained in terms of a still emerging theory of self-stimulation or feedback.

EXPERIMENT II

The above interpretation of the results in Experiment I assumes that the pseudo-words having high spelling-to-sound correspondences are actually perceived (read) more quickly and accurately than those with low correspondence. Proponents of theories emphasizing emitted frequency or 'response-bias' will be tempted to object that the difference between the two sets of words was not due to a difference in perceptibility, but instead due to the S's tendency to emit, whatever the stimulus presented, the responses which were in the past made most frequently.⁴⁰

The 'spew hypothesis' of Underwood and Schultz⁴¹ states the operation of response-bias as follows: "The order of emission of verbal units is directly related to frequency of experience with those units." Are the pronounceable words recorded correctly more

⁴⁰. See, for example, Israel Goldiamond and W. F. Hawkins, *Vexiersversuch: The log relationship between word-frequency and recognition obtained in the absence of stimulus words*, J. exp. Psychol., 56, 1958, 457-463; C. R. Brown and Herbert Rubenstein, *Test of response bias explanation of word-frequency effect*, Science, 133, 1961, 280-281.

⁴¹. B. J. Underwood and R. W. Schulz, *Meaningfulness and Verbal Learning*, 1960, 86.

often by the Ss only because they are nearer the threshold of availability as responses, or are they actually perceived better?

If emitted frequency, in the past, is responsible for the words recorded, one might expect that errors in the previous experiments would include many which changed spelling without changing pronunciation. An example would be writing KLERPHT for KLERFT or GLURK for GLURCK. Some errors of this type were made, but fewer than seem consistent with a serious consideration of the hypothesis. Only 16 errors (out of a total of 7250 opportunities to respond) could be classified as identical in pronunciation with the pseudo-word exposed on the screen. But there are so many possible causes of error that one has no expected frequency with which to compare this number. A more direct test therefore seemed advisable.

If the S is not required to emit a word-response at all, but is instead asked to recognize or match one word with another, we can be fairly sure we are testing what he has perceived, not merely the response which is most available and easy to emit. He does not have to produce the response at all, in this case. One could use a judgment of same or different, or a match from a list including the word shown and some incorrect versions.

Our second experiment was designed to test the hypothesis that an S would recognize the word shown, from a multiple-choice list including this word, more accurately when the word shown had a high spelling-to-sound correlation.

Construction of lists. The pseudo-words exposed for recognition were 20 from the two lists (pronounceable and unpronounceable)

used in Experiment I. The 'misleads' for the multiple-choice series for each one posed a problem. Artificial construction of these by made-up rules would yield unknown and probably unequal difficulty for different words. Also, the S might discover some of the rules before finishing and thereafter choose on a problem-solving basis, rather than perceptual matching. The series was obtained empirically, therefore, by including for each word the three most frequent errors recorded in the previous experiment, whatever they were. For example: for GLURCK the following words were used: GLURK, GLUCK, GLURGK, and GLURCK; for CKURGL the following were used: CKUGL, CKURL, CKUGEL, and CKURGL. Since one position in the choice-list might be more favorable than another, four lists were so constructed for each word that each choice appeared once in each of the four positions.

Practice-effects were equated by making two random orders of the words to be presented tachistoscopically, the second one so constructed that the unpronounceable version of a pronounceable combination now appeared in the position the pronounceable one had occupied in the first order (and vice versa). There were therefore 8 sub-groups constituted by the two orders, each with four versions of the multiple-choice series.

Procedure. S was seated 5 to 7 feet from a screen on which the words were projected, as before. The width of the total field projected was 31 inches. The width of the word on this field varied from 12 to 24 inches, depending on number of letters. The letters, as projected were 3 1/2 inches high. Ss were run individually. The instructions read to S were as follows:

On the screen before you, between the two black lines, a nonsense-word will appear for only a fraction of a second. Immediately afterward, a list of four nonsense words will appear, one of which will be the word just exposed.

TABLE II

Number of Correct Choices of P and U Words Separated
for Order and Multiple-Choice Matching List

Order	Words	List 1	List 2	List 3	List 4	Mean
A	P	121*	171+	105*	106†	14.37
	U	92*	130+	83*	91‡	11.31
B	P	96§	91§	84§	103†	14.96
	U	63§	74§	68§	88‡	11.72
Mean	P	15.50	14.56	13.50	14.93	14.62
Mean	U	11.07	11.33	10.79	12.79	11.48

*N = 8; +N = 12; †N = 7; §N = 6.

Choose the one you think is the same as the word you have just seen and place a number in the blank corresponding to its order on the screen. Number 1 is at the top and Number 4 at the bottom. A 'ready' signal will be given before the first exposure. When the 'ready' signal is given, please be sure to fixate between the two lines so as not to miss the word.

Do not try to make inferences about the alternatives provided in the choice-list, because they were obtained experimentally. Just look at the word flashed, and pick the one in the list which looks most like it.

Ten practice-words were given first to accustom S to the timing of the exposure. A pseudo-word was exposed tachistoscopically for 30 m. sec. and was followed after a half second by projection of the appropriate choice-list. This list was projected for a duration of

8 seconds. The duration was long enough to permit S to read the words and record one, but not to study the alternatives in detail.

The 40 experimental words followed, in the order assigned to the sub-group. The entire experiment lasted about 18 minutes. There were 60 Ss, all university students. Each sub-group contained at least 6 Ss. An S took part in only one sub-group.

Results. The results were first scored separately for each sub-group, to see whether trends changed depending on order in the whole series or position in the choice-list. Table II contains the results by sub-groups. The mean number correct with a given order can be read across the table. The mean number correct for a given order in the choice-list can be read downward. For every sub-group (Order A combined, Order B combined, and Choice-Lists 1, 2, 3, and 4) the pronounceable words were matched correctly more often than the unpronounceable words. It seemed justifiable, therefore, to combine all the results. The over-all means were 14.62 correct for pronounceable words and 11.48 correct for unpronounceable words.

Since a quarter of the choices would be expected to be correct by chance, the number perceived correctly beyond chance was 9.62 for P words and 6.48 for U words. This is a gain of 48 percent for P words over U words.

When the words (pronounceable and unpronounceable) were matched by pairs, 15 of the 20 pronounceable words were perceived correctly more often than their unpronounceable mates. In four cases, the unpronounceable version was perceived correctly more often, and in one case there was no difference. By the Wilcoxon Matched-Pairs

Signed-Rank Test (two-tailed) the difference between the lists was significant at $p < 0.002$. The differences were also significant ($p < 0.01$) when tested by Ss.

The results of this experiment, then, corroborate those of the previous one and provide further positive evidence that pseudo-words constructed according to rules of invariant spelling-to-sound correlation of the English language are perceived more accurately under conditions of tachistoscopic viewing than their matches with no invariant spelling-to-sound prediction. It seems reasonable to conclude that the perception of the stimulus-word is affected by this correspondence; that we are not observing simply a 'response-bias' to emit pronounceable combinations.

When responses must be emitted, however (as in Experiment I), there might conceivably be a combined effect of the perceptual bias toward grasping the higher-order invariant and a response-bias toward pronunciability. In that case, the difference between the two word-lists in Experiment I should be considerably greater than that in Experiment II. It is impossible to make a direct comparison here, since a longer list (25 pairs) was used in Experiment I, as well as a different method (reproduction vs. recognition). The advantage of the pronounceable words over the unpronounceable words at the first exposure (30 m. sec.) was 65 percent in Experiment I, as against 48 percent in Experiment II. It is possible, therefore, that both effects are operative in Experiment I, resulting in greater difference in the results.

The words in the P and U lists.

The question now to be answered is how the rules of invariance for grapheme-phoneme correspondence are learned. To answer this question it may be feasible to 'build in' such units experimentally. If this can be done, the learning process presumably could be studied. Meanwhile, a simple associational model does not seem the best to entertain, for the rules of correspondence are conditional (whether a cluster is at the beginning or the end or the middle, etc.). A trigram such as 'ign' functions as a predictable cluster for pronunciation at the end of a word (e.g., sign, malign) but the prediction will be different in another context (e.g., significant). A model more like a paradigm for concept-formation, with conditional isolation from context, seems more promising. Hopefully, a workable model can be formulated. Developmental experiments which may reveal the presence or absence of functional units of the kind posited at different age-levels are now in progress, and should provide some clues.

CONTROL EXPERIMENTS ON PRONUNCIABILITY

Rating of pronunciability. Are letter-combinations which have, according to the rules, an invariant grapheme-phoneme correspondence, actually more 'pronounceable'? The term 'pronunciability' has been used in this paper as if such a statement could be made. But the matter is not merely one of convenience of reference, since Underwood and Schulz found that pronunciability, as defined by their scaling procedure, was a very important variable in verbal learning.⁴² Ratings were therefore obtained of the pronunciability of the words in the P and U lists.

⁴². Underwood and Schulz, op. cit., 163-199.

The rating method devised by Underwood and Shulz was copied verbatim except for references to word-length and 'common' words (some of theirs were common words, while all of ours were pseudo-words). The two sets of words were mixed and presented to a group of 165 Ss (students in an introductory class in psychology) for rating on a 9-point scale. A score of 1 equals high and 9 equals low. Order was reversed for half the Ss in case scale changes would occur with practice.

The results showed a clear difference in pronunciability between the two lists. Mean ratings were computed for each of the 50 pseudo-words (see Table I). The over-all mean rating of the pronounceable words was 2.88 (SD = 1.48); that of the unpronounceable words was 6.57 (SD = 0.94). For every one of the matched pairs the pronounceable combination received a lower mean rating (i.e., more pronounceable) than the unpronounceable combination. In all cases the mean for a pronounceable combination was lower than the over-all mean for the unpronounceable ones; similarly, the rating of each unpronounceable combination was higher than the over-all mean of the pronounceable combinations. There was considerable inter-individual variability in rating, probably related to differences in background of language, but the results fully justify the characterization of one list as more pronounceable than the other.

Invariance of pronunciation. A second question may be asked regarding the actual pronunciation of the pseudo-words. The rules of grapheme-phoneme correspondence imply that there should be a relatively invariant pronunciation of the P words, but a variety of pronunciations for the U words, which do not follow the rules.

Is this indeed a fact? To answer the question, 16 Ss were asked to read the 50 words aloud. They first read quickly aloud through the two lists, one and then the other. The total time was noted for each list. Then they read the words, mixed in a random order, with a 5-second pause between words. The words were exposed, one at a time, on cards. A tape recording was made for each S. Order was balanced among the Ss.

The first reading served both as practice and as a time-check. If readability under conditions of tachistoscopic exposure is related to reading skill under normal conditions, the total reading time should differ for the two lists. The mean reading time was 38.8 seconds for the P list and 57.5 seconds for the U list, thus confirming this expectation. The trend was consistent for individual Ss.

The tape recordings (all 50 words for each of the 16 Ss) were given to two linguists for analysis of the pronunciations. They worked individually. The method used on their first transcriptions was partly phonetic and partly phonemic. The sequences of letters in the U list, and perhaps occasionally in the P list, were such that the Ss sometimes brought in phonemes from other languages to make them easier to pronounce. In general, however, the Ss included extra sounds, so as to push the sequence toward a possible English one. On their second analysis, the linguists tried to put all the phonetic transcriptions in terms of the phonemic system of English (that is, in terms of significant contrasts). When the two linguists had disagreed on a transcription, the word was played several times and usually an agreement resulted. There was finally 98 percent agreement between them.

There is very little overlapping between the distribution of different pronunciations for the P list and the U list, although two P words (DRIGHK and SMAMP) were quite inconsistently pronounced and fell within the U distribution. The mean of different pronunciations was 2.68 for the P words and 9.12 for the U words. When P words were compared with their U counterparts, all P words had fewer different pronunciations except one (DRIGHK) which had a number equal to its counterpart.

The words in the U list were invariably pronounced as bisyllabic words, whereas the words in the P list were almost always pronounced as monosyllables.

It can be concluded that pseudo-words constructed according to rules of grapheme-phoneme correspondence are pronounced more consistently when read aloud than those which are not.

Relation between scaled pronunciability and invariance of pronunciation. Whether pronunciability, as measured by a rating scale, is related to invariance of pronunciation (relative to rules of grapheme-phoneme correspondence) can be determined by comparing the results of the two experiments above. The results are independent, since none of the Ss took part in both experiments.

Both sets of scores (ratings and variability of pronunciation) for the 50 words were transformed into standard-scores and correlated by means of a Pearson r . A positive correlation of 0.85 was found (not corrected for attenuation). The magnitude of this correlation is very interesting, for it suggests that the rules of grapheme-phoneme correspondence, which exist objectively in the

language, are reflected psychologically in individual ratings of pronunciability of words and also in group conformity of pronunciation.

These two experiments seem to us to provide quite impressive empirical validation of the rules used in constructing the lists.

SUMMARY AND CONCLUSIONS

The hypothesis has been proposed that the proper unit of the reading process is neither the single letter nor the whole word but a higher-order invariant derived from grapheme-phoneme correspondences. Rules can be drawn up for predicting spelling from sound if temporal patterning of 'clusters' is taken into account. It was proposed that the skilled reader has learned to perceive as units the letter-patterns conforming to these rules and that such patterns have, therefore, an advantage in reading.

Several experiments were performed to test this hypothesis. Pseudo-words conforming to rules of spelling-to-sound correlation were compared with ones of low correlation under conditions of tachistoscopic viewing.

In the first experiment, it was found that the Ss reproduce pseudo-words of high correlation more accurately than ones matched for summed letter-frequency, but of low correlation.

In a second experiment, comparable results were obtained when the Ss were not asked to reproduce a word, but performed perceptual matching of the word exposed tachistoscopically with a member of a multiple-choice list. It was concluded that the perceptual process has been facilitated in skilled readers for units discovered during long exposures to the grapheme-phoneme correspondences of the English language.

A STUDY OF THE DEVELOPMENT OF GRAPHEME-PHONEME

CORRESPONDENCES¹

Eleanor J. Gibson, Harry Osser,
and Ann D. Pick

Cornell University, Ithaca, New York

The advocates of spelling reform have long been arguing that the English language affords only a low predictive correlation from spelling to sound (grapheme to phoneme). A written letter is, essentially, an instruction to utter a phonemic unit characteristic of the language; but the 26 letters must somehow map into, or combine to specify 38 or more different English phonemes (Hockett, 1958, p. 59 ff.). How is it possible, then, that a child can learn to read except by forming an association between each written word as a whole and the auditory-vocal word? Fortunately, such an uneconomical learning procedure as the "whole-word" method is not the only possible one. Hockett (1960) and Venezky (1962) have demonstrated that grapheme-phoneme correspondences have high predictability, even in English, if rules are sought in higher-order units rather than in the single letters. The rules are framed in terms of vowel and consonant spellings instead of single letters (e.g., in the word weight, "eigh" might be considered a vowel spelling for mapping purposes). Position contingencies of spellings are also important for prediction of pronunciation, i.e., what precedes and what follows a vowel spelling, or whether a consonant cluster is at the beginning or end of a word. The consonant cluster GH is sounded differently at the beginning than at the end of a word (e.g., as in *ghost* and *enough*). Gibson, Pick, Osser, and Hammond (in press) have

1. To be published in Journal of Verbal Learning and Verbal Behavior.

demonstrated, furthermore, that the letter-sequences generated by these rules function as units in perception, even when they are meaningless nonsense sequences. The rules are, presumably, inferred from real words already encountered and thereafter transfer to the perception of new words and also to the perception of pseudo-words, irrespective of meaning.

To summarize, the hypothesis asserts that higher-order rules exist which predict sound from spelling and that these are assimilated by the skilled reader with the effects of organizing larger units for perception and predicting the way to read hitherto unseen words. How is this different from a frequency hypothesis? Are not certain letter combinations, being more frequent, read with greater speed or at a lower threshold than others? It is true that for discovery of a correlation or an invariant relationship between graphic and phonic patterns, there has to be frequency of something. But frequency of what? Neither summed letter frequency (Gibson *et al.*, in press) or trigram frequency (Postman and Conger, 1954) are good predictors of lowered thresholds in the recognition of pseudo-words. Frequency counts will be relevant only if the right units and relationships are counted. Neither individual letters nor arbitrary groupings of letters, such as diagrams or trigrams, are relevant. The relevant graphic unit is a functional group of one or more letters, in a given position within the written word, which is in correspondence with a specified pronunciation. A frequency count of initial "G" would be pointless. But a count of initial "G" before a given vowel or vowel cluster and representing a given sound would be relevant. Frequency counts within this framework of contingencies should be available when the computer program (Venezky, 1962) has been run.

"Pronunciability" has been used as a shortcut term to describe letter sequences which are generated by the rules of grapheme-phoneme correspondence, as opposed to those which are not. This term might have a certain ambiguity for our purposes. It may mean that a given combination of letters rolls off the tongue easily. Underwood and Shulz (1960) probably intended this meaning. But it also has the meaning of invariance, or relative invariance of the way of pronouncing a given letter sequence. These two meanings were operationally defined and the letter groups of the earlier experiment (Gibson *et al.*, in press) scaled in both ways. The measurements for the two scales correlated better than $+0.85$. We feel justified, therefore, in using pronunciability ratings (and either meaning of the word) when the labor of gathering the data for variability of pronunciations and their phonetic transcriptions is too expensive. Ease of pronouncing a letter-group, in other words, has been shown to be a good indicator of relative invariance and also of "packaging" or unit-formation in perception.

The question examined in this paper is how the correspondence rules are learned. Two quite different possibilities suggest themselves. The obvious one, which fits the usual educational procedure, is that the child begins by memorizing whole words. He learns by rote, with flash cards and so on, to associate the printed CAT with the equivalent spoken word. Later, perhaps in second or third grade, his teacher introduces some phonics, and teaches him how to "sound out" and analyze a word. (This procedure is called "word attack"). Now he might begin to formulate some of the correspondence rules. On the other hand, the correspondence rules might develop in the way that grammar appears to develop (though we have a great deal to learn about this process). A

child speaks grammatically nearly as soon as he speaks at all, even though the sequence is short and the grammar very simple (see, for example, Brown and Fraser, in press). Perhaps as soon as a child gets the idea that writing and printing stand for the sounds of his spoken language, he begins to induce for himself the predictive regularities, even though the words are short and the rules, therefore, quite simple ones. One would not expect that he would formulate the rules verbally, any more than a child of four does the rules of grammar.

If the first alternative is the way the average child learns, it should follow that children in the first grade would read speedily and as units the words which have appeared frequently in their reading drill, but not pseudo-words which have never been encountered, even though they fit the correspondence rules (i.e., have a predictable, invariant pronunciation). Third graders should have progressed to the stage of reading efficiently pseudo-words as well, when they fit simple correspondence rules. But if the second alternative is correct, it would be expected that even first graders, by the end of a year, would read pronounceable pseudo-words more efficiently than unpronounceable ones. Progress of the third graders should be exhibited in increased skill of reading pseudo-words generated by more complex rules (more elaborate organization, with sub-units, for instance).

Following this reasoning, an experiment was designed to compare children at the end of Grade 1 and at the end of Grade 3 in ability to recognize familiar three-letter words, pronounceable trigrams and unpronounceable trigrams. By either hypothesis, the familiar words should be easy to recognize; the question was whether there would be a difference between the two types of trigram and, if so, whether this

difference would already be present by the end of the first grade or would develop only later. Some longer pseudo-words were included as well, to investigate the span (word-length and complexity) over which any rules might operate.

Words from the first-grade reading list were chosen which could be broken into meaningless trigrams, both a pronounceable one and an unpronounceable one from each word. Trigram frequencies were then tabulated and lists chosen so that trigram frequency was as great or greater overall for the unpronounceable as for the pronounceable items. The frequencies were taken from Underwood and Schulz's (1960) tables (the total column for trigrams which combines three separate counts). Table 1 shows the trigrams and words.

TABLE 1
Three-Letter Words, Pronounceable Trigrams, and
Unpronounceable Trigrams and Their Frequencies

Word	f	P trigram	f	U trigram	
PUT	40	TUP	0	PTU	15
AND	1070	NAD	6	DNA	4
PIG	44	GIP	0	GPI	2
SUN	131	NUS	7	NSU	46
MAN	395	NAM	8	NMA	12
CAN	216	NAC	7	NCA	14
PET	40	TEP	13	PTE	15
CAT	249	TAC	19	TCA	1
RAW	269	NAR	26	RRA	27
GUN	5	NUG	10	NGU	13
Total	2459		96		149

The longer pseudo-words were 20 four- and five-letter combinations used in the previous experiment with adults (Gibson *et al.*), half pronounceable and half unpronounceable. Each pronounceable word had

an unpronounceable counterpart, as before, constructed by reversal of the initial and final consonant spellings.

METHOD

Instructions.

The trigrams and words were exposed tachistoscopically one at a time to an individual subject. The instructions were as follows:

"I'm going to flash some letters on the screen in front of you, very fast. The game is to see if you can read them to me. Sometimes they will make a real word, sometimes not. Read them like a word and then spell them to me. I'll say 'ready,' and then you watch the screen for the flash. Ready . . . What did it say?"

These instructions included the admonition to spell out the letters, because it was found in preliminary experimentation that the children usually refused to say anything when highly unpronounceable combinations were presented. Yet, with probing, it turned out that the child had sometimes seen something. In order not to "stack" the data in favor of the hypothesis, it was thought necessary, therefore, to include the spelling instruction. Fairly consistently, the children pronounced any pronounceable combinations as a unit and then spelled it, but merely spelled the others. The spelling was recorded as the response to be treated.

Procedure

The child was given practice words and all of his questions were answered. The trigrams and three-letter words appeared first in random order (two random orders were used with positions of pronounceable and unpronounceable items exchanged). Before the longer words appeared, the

child was told that they would be longer than the words he had seen previously.

Exposure time for the words was constant at .04 sec. Shorter exposure times were explored, but proved impossible for the first-graders. Contrast (figure and ground) was higher than that used in previous experiments with adults (line brightness was approximately $7/8$ of field brightness). The illumination of the pre-exposure field was slightly higher than the exposure field (2.6 ft. lamberts as against 1.7 ft. lamberts). The settings described were maintained at a constant level for all children throughout the experiment.

Subjects

There were 24 subjects (12 boys and 12 girls) who had just completed 1st grade and as many who had completed 3rd grade. They were drawn from the same school population,² with names picked at random. A letter of explanation was sent to parents of prospective Ss and a telephone call followed. Only one parent refused. The children were called for one or two at a time by the E and brought to the laboratory, so there was ample time to make them feel at ease. The children seemed to enjoy the experiment, although one child said, when called for, "I won't come if it's about Dick and Jane."

RESULTS

Table 2 presents the results for the three-letter words and trigrams, divided according to grade and sex. The mean is number read correctly out of ten. Only a whole word or trigram correct was counted here.

2. The school was the one nearest the University, so a number of the children came from homes of higher than average educational background.

TABLE 2
Means and Percentages of Three-Letter Words and
Trigrams Read Correctly

		Grade 1		Grade 3	
		Boys	Girls	Boys	Girls
Words	M	6.58	6.42	6.83	7.83
	%	65.8	64.2	68.3	78.3
Pronounceable trigrams	M	4.25	5.42	5.50	7.92
	%	42.5	54.2	55.0	79.2
Unpronounceable trigrams	M	3.58	4.25	4.83	7.5
	%	35.8	42.5	48.3	75.0

(Part scoring was difficult and showed no different trends.) Others were compared first. Since they were not significantly different, they were combined in all later analyses.

A three-way analysis of variance was performed on the trigram data. The analysis showed a significant treatments effect: words were read correctly with greatest frequency, and pronounceable trigrams surpassed the unpronounceable ones. The grade effect was also significant. The only significant interaction was Treatments X Sex. This interaction was caused by the 3rd grade girls who differed from the other groups in showing no difference among the three types of items. Their accuracy was 75 percent or better for all three, which is so high as to be near the limit for the method of scoring by whole items. In this respect, they behave like adults with three-letter items under these exposure conditions--all are ready with nearly equal accuracy.

The four- and five letter pseudo-words were not included in the analysis of variance. Results for these items are presented in Table 3.

TABLE 3

Means and Percentages of Four- and Five-Letter Pseudo-Words Read Correctly

		Grade 1	Grade 3
Pronounceable	Mean	.46	3.50
	%	4.6	35
Unpronounceable	Mean	.63	1.88
	%	6.3	18.8

First graders were able to read very few of these words; a mean of less than one from each list (n of 10 per list). The difference between the pronounceable and the unpronounceable items is not significant ($P > .388$ by a sign test, two-tailed). But the third graders have an increased span; furthermore, the difference between the pronounceable and the unpronounceable items is significant ($P < .004$ by a sign test, two-tailed). The proportion of pronounceable to unpronounceable items is also different for the two grades (18:1 for third grade, 4:8 for first).

SUMMARY AND CONCLUSIONS

The results of this study can be summarized as follows.

(1) Familiar three-letter words which had been practiced as units were read more accurately by all the first graders tested and by the third-grade boys than were non-word trigrams.

(2) Pronounceable trigrams were read more accurately than unpronounceable ones by the same children, with both letter and trigram frequency controlled.

(3) Third-grade girls read all three-letter combinations with high and equal accuracy, resembling adult readers.

(4) Span (length of word which can be read under these conditions) increased from first to third grade.

(5) As span increased, a difference in reading accuracy also developed between longer pronounceable and unpronounceable pseudo-words.

(6) Tentatively, it is concluded that a child in the early stages of reading skill typically reads in short units but has already generalized certain consistent predictions of grapheme-phoneme correspondence, so that units which fit these simple "rules" are more easily read. As skill develops, span increases and a similar difference can be observed for longer spans. The longer items involve more complex contingency rules and longer vowel and consonant spellings, so that the generalizations must increase in complexity as the span increases. Thus, the second hypothesis suggested in the introduction is supported. Even though a child is presented with "whole words" and encouraged to associate the printed word as a whole with the spoken word, he still begins to perceive some regularities of correspondence between the printed and written terms and transfers these to the reading of unfamiliar items. This generalizing process undoubtedly promotes reading efficiency and could be facilitated by presenting material in such a way as to enhance the regularities and speed up their incorporation.

References

- Brown, R., and Fraser, C. The acquisition of syntax. In Cofer, C. (Ed.), Verbal behavior and learning: Problems and process. New York: McGraw-Hill, in press.
- Gibson, Eleanor, Pick, Ann, Osser, Harry, and Hammond, Marcia. The role of grapheme-phoneme correspondence in the perception of words. Amer. J. Psychol., in press.
- Hockett, C. F. A course in modern linguistics. New York: Macmillan, 1958.
- Hockett, C. F. Analysis of English spelling. Mimeographed paper, June, 1960.
- Postman, L., and Conger, B. Verbal habits and the visual recognition of words. Science, 1954, 119, 671-673.
- Underwood, B. J., and Schulz, R. W. Meaningfulness and verbal learning. Philadelphia: Lippincott, 1960.
- Venezky, Richard. A computer program for deriving spelling-to-sound correlations. Unpublished Master's thesis, Cornell Univ., Ithaca, New York.

A COMPARISON OF MEANINGFULNESS AND PRONUNCIABILITY AS GROUPING
PRINCIPLES IN THE PERCEPTION AND RETENTION OF VERBAL MATERIAL¹

Eleanor J. Gibson, Carol H. Bishop, William Schiff and Jesse Smith

Cornell University

Hardly anyone would question the proposition that the efficient reader or thinker deals with larger units of material, whatever they may be, than the beginning or retarded practitioner. It was not necessary for psychologists to point out that speedy readers do not dwell on every letter, but take in words, phrases and conceivably even sentences "at a glance." Unlike some common-sense knowledge, however, this maxim has received convincing corroboration from well-conducted experiments. Perhaps the most famous of these is Bryan and Harter's (1899) which demonstrated that progress beyond a certain point in the reception of Morse code depended on the development of higher-order units. The importance of higher units in the perception of verbal material was early demonstrated by Cattell (1885) and by Erdmann and Dodge (1898). The number of letters which can be perceived in a tachistoscopic exposure increases as the letters form a word, and again as the words form sentences.

For several decades psychologists accepted these facts with satisfaction and with little further curiosity. But in recent years, new interest has centered on the problem of how the higher units are formed. At least three lines of investigation have led to this revival. One is the advent of information theory and the realization that information processing is affected by what Miller (1956) called "chunking." Another is the emphasis of rote-learning theorists on "response integration" (Mandler, 1954, 1962; Underwood & Schulz, 1960); and third, experiments

1. Submitted to J. Exp. Psychol.

on the reading process considered as perceptual learning (Gibson, Pick, Osser & Hammond, 1962; Gibson, Osser & Pick, 1963).

The general questions to be answered are several. How are higher order units formed? In fact, are they formed by the learner, or are they rather discovered in the structure of the stimulus array? What principles of grouping, in either case, favor the development of higher units, and are these principles the same, whatever the task to be performed with the units?

It was generally taken for granted in the early work on tachistoscopic perception of word vs. non-word sequences of letters that words were read faster because they were "meaningful"; meaning presumably conferred unity on the letter group. But recent work (Underwood & Schulz, 1960; Gibson et al., 1962) makes plausible the hypothesis that pronunciability -- the correlation of a spelling pattern with an invariant pronunciation -- is an alternative and important grouping principle for letter sequences, a cohesive property making for unit formation.

Is it possible to separate these two properties in order to compare their effectiveness? The difficulty of defining meaningfulness (or at least agreeing on a definition satisfactory to more than a few people) is at first discouraging. Furthermore, real words, however meaningful by any criterion, are de facto pronounceable. And nonsense syllables, when rated for meaningfulness and for pronunciability, appear to have these properties correlated. This fact is not surprising, for pronunciability is itself a kind of meaning. A pronounceable letter group obeys rules of English grapheme-phoneme correspondence and therefore has a structure which is familiar and related to other verbal structures. But it may nevertheless lack semantic reference and thus be low in that aspect of meaning.

A method of separating meaning in the sense of semantic reference from pronunciability was suggested to us by the experiments of Underwood and Schulz (1960). They found that pronunciability of trigrams was an excellent predictor of learning, but there were some notable exceptions. These were letter sequences which had either extremely high generated value (i.e., xyz), or very consistent and familiar semantic reference (i.e., initials such as RFD). These two indices of meaning are quite different from the one implied by Noble's m (an index of meaningfulness given "by the average number of associations made by a representative sample of subjects during a standard time interval," Noble, 1963, p. 84), but they can nevertheless be defended as equally relevant. Underwood and Schulz considered generated value of the kind found in xyz a useful measure of meaningfulness. More specifically, they said "We are interpreting both pronunciability and the generated values to be measures of response integration" (p. 246). As for consistent semantic reference, it may be that a letter group is no less meaningful when it has one and only one association which everyone agrees on than when it has a numerous and varied set of them.

Familiar initials satisfy both these criteria of meaningfulness; they have high generated frequency of the third letter, given the first two, and they have consistent semantic reference. Furthermore, a number of them defy pronunciation as a unit, so that the letter sequence can be considered high in one property and low in the other. Underwood and Schulz said, in comparing generated values with pronunciability, "Thus if we capitalize on alphabetical-sequence habits, initial habits, and so on, which often result in trigrams difficult to pronounce, we could quite easily reverse the relative predictive capacity shown for the two measures in the present experiment" (p.258).

Whether there are differences in the effectiveness of verbal units "packaged" primarily by meaningfulness as just defined, or by pronunciability is the subject of this investigation. Their effectiveness as grouping principles is compared by determining perceptual thresholds for items high in either pronunciability or in meaningfulness, but low in the other, and for control items low in both. Their effectiveness is also compared for two measures of retention.

METHOD

Material.--Three types of trigrams were selected for comparison: trigrams whose unitariness was constituted primarily by their being pronounceable as a monosyllabic unit, according to the rules of English spelling and pronunciation; trigrams whose unitariness was constituted primarily by meaningfulness defined by semantic reference and generated value; and control trigrams low in both pronunciability and meaningfulness.

Since we considered it important to control the letter-shapes, which are themselves a source of error in measuring perceptual thresholds, the trigrams were chosen so that each control trigram, rearranged anagram-wise, yielded a pronounceable counterpart and a meaningful counterpart. For example, a control trigram MBI had as its pronounceable counterpart MIB and as its meaningful counterpart IBM.

First a search was made for trigrams yielding a triad of the kind described. Next, all these items were checked for generated frequency of the third letter in the tables in Underwood and Schulz (p. 377 ff.), since it was desired that meaningful items be high in this respect and the others low. When there was a choice between two control or two

pronounceable trigrams (e.g., MIB or BIM), the one with lower generated frequency of the third letter was chosen.

After these steps, we were left with 15 sets of trigrams to be rated for meaningfulness by a population similar to that finally to be used as Ss. Subordinate lists of 15 trigrams were constructed, each containing only one member of a triad such as IBM, MIB, and MBI. Two versions of each of the three resulting lists were made with different random orders. Five real words (the same ones) were inserted in each of the lists as well. Because the first try-out indicated that some of our would-be meaningful items might turn out to be low for the population tested and would have to be discarded, four more such items were added to all the lists to be rated. Each list finally included nine supposedly meaningful trigrams, five pronounceable trigrams, five control trigrams and five words.

The six subordinate lists were distributed to different Freshman English classes. Each item was rated for meaningfulness on a 4-point scale (very meaningful, somewhat meaningful, slightly meaningful, not at all meaningful). In addition, the rater was asked to indicate the meaning of each item, if it had any, in order to locate any idiosyncratic items. Each trigram was rated by at least 30 to 40 people.

Table 1 shows the 12 sets of trigrams finally chosen, with the percentage of Ss that rated the trigrams very meaningful or somewhat meaningful. Of the 19 items intended to be meaningful, seven had to be discarded, either because they were not rated high enough by a majority of Ss, or because one of the two counterparts was rated high. Two of those retained were ones inserted after the first

TABLE 1

Per Cent of Subjects Rating Trigrams as Very Meaningful
or Somewhat Meaningful

Control		Pronounceable		Meaningful	
Trigram	%	Trigram	%	Trigram	%
OKR	9	KOR	5	RKO	74
AVT	9	TAV	6	TWA	88
MBI	3	MIB	12	IBM	100
IFB	7	BIF	11	FBI	100
ACB	12	BAC	16	ABC	95
ATW	0	TAW	3	TWA	93
RLI	0	RIL	8	ILR	95
APW	3	WAP	24	WPA	71
OTK	4	TOK	3	TKO	91
YNU	7	YUN	6	NYU	94
ICF	-	CIF	-	IFC	95
LFA	-	FAL	-	AFL	93

tryout and did not, therefore, have their counterparts rated. The meanings indicated by Ss were examined for idiosyncrasy. Items retained as meaningful had consistent reference, high generated value, and a high meaningfulness rating.

Ratings of pronunciability were not obtained. Instead, the trigrams were presented to a group of judges (members of a graduate seminar in verbal learning). There was unanimous agreement that the items in the pronounceable list were pronounceable as a monosyllabic unit according to English usage, and were also easier to pronounce, by a native English speaker, than their two counterparts.

Lists to be used for testing perceptual thresholds and retention were "mixed," that is, made up with four items of each of the three groups. A list contained only one member of a triad. All items appeared in a list an equal number of times and were distributed so as to randomize list membership and so as to counterbalance for possible position effects within a list. Order within the list was random, but with the stipulation that items of the same type be separated by two other items. Lists of items presented for threshold determination included four real words interspersed among the trigrams. Ratings of meaningfulness for these words (sum, hog, rat, joy) were lower on the average than for the meaningful trigrams, suggesting that a successful effort had been made to achieve high meaningfulness for these trigrams.

Method for determining perceptual thresholds.--Thresholds were measured individually for each S by means of varying the brightness contrast of the exposure field in relation to the pre- and post-exposure fields. The time of exposure was maintained constant at .25 sec. Brightness of illumination of the pre- and post-

exposure field was also set at a constant level for all Ss. The brightness of the exposure field was variable, always starting with an illumination well below that of the pre-exposure field and increasing by 1 volt steps up to the perceptual threshold. The tachistoscope was a double projector set-up with synchronized camera shutters. Variacs wired into the projector circuit permitted manipulations of the voltage. S sat in front of the projector at a distance of 8 feet from the screen. The material was typed in Roman capitals with a sign-typewriter and prepared as slides. Height of the letters projected on the screen was 17.5 cm. A slide in the projector for the pre-exposure field contained two black lines so placed as to frame the trigrams to be projected. Illumination of the projection room was low and always constant. Illumination of the pre-exposure field (set at 90 volts) measured 14 foot lamberts. When the background illumination of the exposure field was at 60 volts (a setting which permitted most of the material to be read) it measured 1.6 foot lamberts and a projected letter-stroke measured .64.

The procedure for obtaining the threshold was an ascending method of limits, with S responding whenever he could by writing what he saw on a pad of paper on his chair arm. Each time he wrote something, E observed it and either increased the exposure illumination by 1 volt for another exposure, or told S he had read correctly and turned the page for recording responses to the next stimulus. After a successful trial, E read the voltmeter, reset it for trials with a new trigram, and gave S a ready signal. The instructions were as follows: "This is an experiment on reading. We would like you to look at the

screen in front of you, and when we say 'ready' fixate the center of the two black lines. After the ready signal, a group of letters will be flashed on the screen between the lines for a brief interval. Please try to read it and write what you see on your response sheet. After writing, please look away from the screen until the next 'ready' signal. At first, you will probably be unable to read anything. In that case, write nothing. We will continue with successive exposures until you have read a letter group correctly. As soon as you are fairly sure you have seen something, write it. But don't merely guess -- write what you have seen only if you have a fair degree of confidence."

Three practice slides were given S (two 3-digit numbers and a word, BOB) during which the voltage suitable as a starting level for a given S was ascertained. The level was set so that S would be unable to report anything on the first exposure, but an attempt was made not to set it so low that any threshold measure would require more than about 15 steps. The 16 trigrams and words were then presented to S. Forty-two Ss were run, yielding 14 threshold measures for each of the 36 trigrams. All Ss were drawn from an introductory psychology course in the College of Arts and Sciences. None had rated the trigrams.

Method for testing recognition-memory.--The same three sets of trigrams were presented in mixed lists of 12, again distributed so that only one member of a triad appeared in a list, but all appeared with equal frequency. As in the previous experiment, there were four trigrams of each type in a list. No real words were included. The three lists thus constituted were arranged in three orders with items rotated in

each order so as to yield identical serial positions across the three members of a triad. The purpose of the three orders was to increase generality of any trend found, since counterbalancing for position was accomplished by maintaining identical positions for members of a triad.

The items were typed with the sign typewriter and photographed on film-strip. A list was projected on a screen with a two-second presentation for each item. Instructions were:

"We are going to show you 12 three-letter combinations on this screen. Please notice them carefully."

After the list had been exposed, S turned over a sheet of paper lying before him and read his further instructions, which were to read through all the letter-combinations printed on the page, and check the 12 he had just seen. This page contained all the 36 trigrams, in random order. Three orders were again prepared so that position in the recognition test could be counterbalanced by assigning members of a triad to identical places. Each of the three test orders was used three times, a given test-order always following a given exposure order so that no bias due to interaction of position of items in exposure list and test list was possible. The list-cum-test combinations were assigned at random to the nine sub-groups, which were sections of an introductory psychology course. The Ns for the counterpart-order sub-groups were approximately equal. None of the 182 Ss had taken part in the previous experiments.

Method for testing recall.--The film-strips made for testing recognition were used for testing recall. However, the recall groups

had three repetitions of a list. A different order appeared for each trial. As before, orders were held constant for members of a triad. At the end of each exposure of the list, S wrote all the items he could recall (in any order) on a pad of paper, turning back the page when he had finished. Again, the lists were projected with a two-second presentation per item and 110 new Ss were run in sections of the introductory course. Each of the three lists was given to two sections, randomly assigned.

RESULTS

Comparison of perceptual thresholds.--The threshold measures which had been recorded in volts were the settings required for an S to read a trigram correctly. As is the usual case with tachistoscopic recognition data, individual differences were very great. In order to determine whether there were further differential effects due to trigram type, the data were transformed to standard scores by the formula $50 + [10 \frac{(x - \bar{x})}{\sigma}]$. The σ was determined for each S individually. Since each member of a triad was judged by different Ss, the conversion to standard scores allowed statistical treatment of the differences between scores in individual triads, eliminating variability due to differences in legibility of letters across triads.

Although this procedure may have allowed slight differences in intra-subject list composition to be obscured, such an effect would tend to work for the null hypothesis instead of against it; and since each S received four trigrams of each type (as well as four words) in his list, no one type of trigram would be favored by the standard score conversion.

The mean thresholds for each trigram, expressed as the converted scores, are given in Table 2. Comparison of the means at the bottom of the columns reveals that the overall mean threshold was lowest for the pronounceable trigrams, next lowest for the meaningful trigrams, and highest for the control items. The mean threshold for each trigram was treated as a score, and paired t tests between item types were calculated, using 11 df for each test. Treating the data in this manner was more conservative than other possible treatments, but the t tests (two-tailed) were significant as follows: pronounceable vs. control, $p < .001$; pronounceable vs. meaningful, $p < .01$; meaningful vs. control, $p = .05$. In fact, the two larger differences would have been significant with a sign test. Eleven of the pronounceable trigrams had lower thresholds than their meaningful counterparts, and all 12 pronounceable ones were lower than their control counterparts. Seven of the meaningful items had lower thresholds than their control counterparts.

It seems clear that pronunciability was more powerful in creating a unit of a three-letter group than was meaningfulness as defined in this experiment, if we accept lowness of perceptual threshold as an indicator of relative unitariness (processing as a whole rather than as discrete elements).

Four actual words were included among each S's stimulus items. Thresholds for the words, since they were both meaningful and pronounceable, might be expected to be lowest of all the items. Such was the case (mean threshold = 45.0).

Analysis of pre-threshold responses.--The pre-threshold responses of all 42 Ss were examined in an attempt to classify errors. Only

Table 2

Mean Thresholds, Recognition, and Retention (Recall) Scores for Triads of Meaningful, Pronounceable, and Control Trigrams

	Meaningful			Pronounceable			Control		
	Thresh- old	% Recog.	% Recall	Thresh- old	% Recog.	% Recall	Thresh- old	% Recog.	% Recall
RKO	45.2	94	68	KOR	64	57	OKR	61	46
TVA	47.2	92	76	TAV	74	18	AVT	51	22
IBM	45.8	95	51	MTB	74	25	MBI	56	22
FBI	57.6	97	73	BIF	92	49	IFB	61	43
ABC	48.3	98	83	BAC	50	51	ACB	80	46
TWA	52.6	95	83	TAW	42	44	ATW	39	11
ILR	52.0	97	83	RIL	71	39	RLI	62	15
WPA	57.9	84	43	WAP	80	51	APW	53	20
TKO	45.7	95	64	TOK	78	61	OTK	63	10
NYU	47.8	100	61	YUN	92	66	YNU	79	49
IFC	61.8	92	86	CIF	71	44	ICF	48	37
AFL	58.2	85	83	FAL	87	50	LFA	46	24
Mean	51.7	94	71.2		73	46.3		58	28.8
SD	5.6								6.4

three-letter responses were included. Two judges placed them in four categories: real words or names, initials or abbreviations, pronounceable trigrams, and trigrams which were not pronounceable and had no known significance. Words or names accounted for 32.5 percent of the errors, 4.7 percent of the errors were initials or abbreviations; 21.5 percent were pronounceable trigrams and the rest were low in pronunciability and meaning. Although the probabilities for the various types of errors to occur are not known, there does not appear to be a set to look for initials during the perceptual task. This inference was corroborated by the Ss when questioned after the experiment.

The contribution of letter-shape confusions to erroneous responses is a matter of considerable importance for reading as well as for control of word-threshold experiments, but it is hard to evaluate quantitatively. One reason for the difficulty is the confounding of letter-similarity and guessing from context. For example, the most frequent pre-threshold response to the trigram BAC was BAG. Is this because BAG is a real word, and S, having perceived the first two letters, is simply operating on the generated frequency principle when he adds the G? Certainly this hypothesis is not entirely satisfactory, for other words such as BAR, BAT, and BAD are highly probable for this population, but were never emitted. On the other hand, C and G are two of the most confusable letters of the alphabet.

In an effort to assess the influence of confusions due to letter-shapes, all the single-letter errors were collected in a confusion matrix. Only those errors which could be unequivocally located in the matrix were counted. (No one or two-letter responses were counted

unless S had inserted dashes to make position in the presented letter-group clear.) The resulting matrix is not, of course, a proper one for studying letter confusability as such, but fortunately a matrix obtained under the most careful conditions with children, using the same type font, was available for comparison. Of the five most frequent letter-confusions found in the pre-threshold responses, four were among the frequent confusions made by the children (C with G, C with O, W with N, M with N). Examples of items yielding a number of these confusions are: ABO for ABC, AOB for ACB, ATN for ATW, NIB for MIB, YMU for YNU. When the structure of the trigram allowed, it seems clear that letter-shape influenced S's erroneous responses as well as his accurate ones.

In analyzing errors, a count was also made of errors in relation to position of the letter in the trigram. Errors were most frequent for the middle letter, slightly less for the third, and least for the first. Greater accuracy for the first letter is to be expected from S's reading habits and from previous research (see Pillsbury, 1897; Heron, 1957). But accuracy on the first letter was by no means perfect. In considering the trigrams which had a higher frequency of errors on the first letter, and comparing them with their counterparts, the influence of letter-shape is detectible. For instance, for the trigram CIF, errors were most on the first letter: for IFC, they were most on the third, and for ICF they were most on the second, with progressively greater frequency in the three cases. There appears, therefore, to be an interaction of position and letter-shape.

Retention measured by recognition.--The percentage of Ss that recognized each trigram correctly may be seen in Table 2. The strong trend of the threshold results, for pronounceable items to have an advantage over meaningful ones, is reversed in the recognition scores. Meaningful trigrams are recognized correctly more often than both pronounceable and control ones (94 percent vs. 73 percent and 58 percent); pronounceable more often than control ones (73 percent vs. 58 percent). All the differences are significant by a sign test ($p < .01$, two-tailed).

The errors of recognition (marking an item as recognized when it was not a member of the list exposed to S) were analyzed to see if any of the three types of trigrams predominated. Most meaningful items were misrecognized (193 errors), control items next (173 errors) and pronounceable ones least (157 errors). Although these differences are not large, the trends were the same in the various subgroups of Ss. Thus, meaning seems to have an advantage for recognition, measured by number of items correctly marked as recognized, but it is also associated with more frequent errors. Pronounceable items are recognized correctly more frequently than control ones, and are less often mis-recognized. The higher frequency of checking the meaningful items, whether correct or incorrect, could therefore be due to a category set (looking for initials) which developed as S saw the list. But the advantage of pronounceable over control items (more chosen correctly and fewer chosen incorrectly) seems more likely due to their being perceived as units at the time of exposure. Structural cohesiveness due to pronunciability as a monosyllable gives S only one item to remember instead of three discrete letters.

Retention measured by recall.--The percentage of items recalled correctly on the first recall trial is presented in Table 2. It was highest for meaningful items, next highest for pronounceable items and lowest for control items. The differences are significant by a sign test (two-tailed) as follows: meaningful vs. pronounceable $p < .04$; meaningful vs. control $p < .001$; pronounceable vs. control $p < .01$. After the first recall trial, the differences between the three types of item decreased. This trend is attributable to the fact that the number of meaningful items recalled correctly did not increase after trial 2.

All the errors (incorrect items written in the recalls) were tabulated and examined for evidence of "coding" by category influence. That category membership (belonging to the class of "initials") is important in recall was suggested by the appearance of items such as YNU in a rearranged form, NYU, in the recalls. Evidence of a category set is also clear in importations such as FDR.

The errors are classified in Table 3. Two judges agreed on the classifications. A test of the differences between observed and expected number of rearrangements to the four categories yields a χ^2 of 65.7 with 3 d.f. This is significant at better than the .001 level.

It is interesting that there were far more rearrangements to initials than to any other category, including real words. There were 44 percent as many opportunities for word errors as for initial errors, but only 3 percent as many errors. The Ss observed, as the list was exposed, that there were no real words but that there were

Table 3

Analysis of Recall Errors

		Observed	Expected
Rearranged to real word	(e.g. CAB)	1	4.6
Imported real words	(e.g. SUP, GNU)	12	
Rearranged to initial category	(e.g. ICF to IFC)	35	10.4
Imported initials	(e.g. CIO)	9	
Rearranged to pronounceable	(e.g. RLI to RIL)	19	27.6
Imported pronounceable	(e.g. TOC)	63	
Rearranged to control type	(e.g. RLI to IRL)	14	26.4
Others		99	

a group of initials, and in recalling tried to make use of this knowledge. Of the 35 cases of rearrangement to the initial category, the item rearranged was more often a control item than a pronounceable item (28 to 7), suggesting that the pronounceable items were more resistant to rearrangement.

There were few rearrangements to pronounceable errors. With 265 percent as many opportunities for pronounceable errors as for initial errors, there were 54 percent as many errors. The pronounceable errors were not many more than were rearranged as the control type. With 105 percent as many opportunities, there were 136 percent as many pronounceable errors as control errors. The category of pronounceable monosyllables is no doubt too large to aid much in recall. Of the errors which were not real words or rearrangements of items exposed, a large number were pronounceable. They were often cases of one letter changed from the item exposed, such as CIV for CIF, TOC

for TOK, and BAK for BAC. It was apparent from some of these errors that S had remembered the sound of a monosyllable and was misspelling it.

DISCUSSION

Two characteristics of verbal units -- pronunciability and meaning -- were compared in these experiments, and both were found to give an advantage to a sequence of letters for reading and remembering. But the striking fact which emerged is the discrepancy between the perceptual results and those for retention as regards the relative effectiveness of the two. Why should pronunciability be more effective here in structuring units for reading but meaningfulness be better for retention? Miller (1962) has pointed out that "language must contain some natural coding units of finite length. But what these units are for an articulate human being still remains something of a mystery" (p. 81). It is possible that the optimal coding units -- or grouping principles -- for one task are not optimal for another. An analysis of S's task in the two situations may help.

In the perception of written language, the perceiver must code the stimulus material into units of spoken language -- that is first and foremost the reading task. But we do not speak in letters; we speak in syllables, words and larger functional units. A letter name is itself a morpheme, not an element of a speech unit. Since the reader must code the written letters into functional speech units, pronounceable letter combinations will obviously facilitate his task. But the kind of meaning present in IBM creates three units for reading rather than one.

In the recognition and recall tasks, on the other hand, the stimulus items have already been read (coded into speech units). The task now is to hang onto them -- store them for future "retrieval." Much recent research has demonstrated that coding for retrieval is most efficient when the items are subsumable under a known and "exhaustible" category (see Cohen, 1963). The category of pronounceable trigrams (if it could be called a category at all) is so large as to be of small help in retrieval. In other words, pronunciability is of the utmost help in structuring a unit -- grouping the letters into a single item -- but it is of little help in grouping those items under a single address. When larger linguistic structures are possible, such as phrases or sentences, they obviously operate to facilitate retention. But linguistic structure can help no farther than the single syllable unit in the present experimental situation.

On the other hand, the class of "well-known initials" does provide a common category for grouping items in storage. It provides a coding principle for retention, and that the Ss used this principle is clear. In recognition, more of the meaningful items were checked than any others, whether right or wrong. And in recall, Ss demonstrated use of the initial category by errors of rearrangement. Underwood and Keppel (1963) found that if an S "encodes" a nonsense syllable into a word and must "decode" it again in recall to the original syllable, performance is inhibited, rather than facilitated. In the present experiment, failure in decoding can be inferred to be responsible for the errors of rearrangement in recognition and recall. The parallel facilitation in both recognition and recall can then be inferred to be the result of grouping the "initial" items

actually presented into one category, from which they were retrieved at the time of recognition or recall.

Consider now the question of response "integration." Meaningful items possessing high "generated frequency" seem to exemplify, maximally, integration by associative frequency. But one cannot conclude that the better retention of these items over pronounceable ones proves the greater effectiveness of associative integration. These items did not have much advantage over control ones as units in perception, and their advantage over pronounceable ones in retention can be accounted for by S's tendency to group them into a useful category. It was thus not only their high letter-sequence associations but the possibility of grouping them under one heading that made the difference. It is interesting to note in this connection that nonsense syllables (paired associates) are harder to learn when they are spelled out than when they are pronounced as units (Rock & Steinfield, 1963). It is possible that initial responses are integrated by associations between the letters but the present results can be explained plausibly without such a principle. That pronounci-ability confers unity is supported.

SUMMARY

Thresholds of visual perception and two measures of retention were obtained for trigrams varying in pronounci-ability and meaningfulness (semantic reference of the kind found in well-known initials) and for control items. The three types of trigrams contained the same letters rearranged anagram-wise. Perceptual thresholds were lowest for pronounceable items, and next lowest for meaningful ones.

On the other hand, retention, measured by both recognition and free recall, was best for the meaningful items and second best for pronounceable ones. Pronunciability was inferred to be the better grouping principle for reading or coding to speech units. Meaningfulness was inferred to have facilitated retention more than pronunciability by providing a category for grouping the initial items, thus aiding retrieval.

Learn. M. Wisniewski, Santa, 1962.

Gibson, E. J., Pick, A. D., Ower, W., & Harwood, H. The role of grapheme-phoneme correspondence in the perception of words. *Am. J. Psychol.*, 1962, 75, 554-570.

Gibson, E. J., Pick, A. D., & Ower, W. A study of the development of grapheme-phoneme correspondence. *J. Verbal Learning and Verbal Behavior*, in press.

Horne, V. Perception as a function of retinal locus and attention. *Am. J. Psychol.*, 1977, 70, 38-43.

Levitt, J. Response factors in mass learning. *Psychol. Rev.*, 1979, 86, 235-264.

Marshall, G. From association to structure. *Psychol. Rev.*, 1962, 69, 425-427.

Miller, G. A. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.*, 1956, 63, 81-97.

Miller, G. A. Decision units in the perception of speech. *All Transactions on Information Theory*, 1962, 10-11, 77-83.

Miller, G. A. Meaningfulness and familiarity. In: *Miller, G. A., Galanter, E. C., & Pribram, K. H., eds. Cognition: The development of learning, problems and processes.* New York: Wiley, 1962.

Pennington, W. B. The reading of words: a study in orthography. *Am. J. Psychol.*, 1907, 9, 313-343.

Rock, I., & Steinfield, G. Methodological questions in the study of non-verbal learning. *Science*, 1963, 140, 1-3.

REFERENCES

- Bryan, W. L., & Harter, N. Studies on the telegraphic language. Psychol. Rev., 1899, 6, 345-375.
- Cattell, J. McK. Ueber die Zeit der Erkennung und Benennung von Schriftzeichen Bildern und Farben. Philos. Stud., 1885, 2, 635-650.
- Cohen, B. H. An investigation of recoding in free recall. J. Exp. Psychol., 1963, in press.
- Erdmann, B., & Dodge, R. Psychologische Untersuchungen über das Lesen. M. Niemeyer, Halle, 1898.
- Gibson, E. J., Pick, A. D., Osser, H., & Hammond, M. The role of grapheme-phoneme.. correspondence in the perception of words. Amer. J. Psychol., 1962, 75, 554-570.
- Gibson, E. J., Pick, A. D., & Osser, H. A study of the development of grapheme-phoneme correspondence. J. Verbal Learning and Verbal Behavior, in press.
- Heron, W. Perception as a function of retinal locus and attention. Amer. J. Psychol., 1957, 70, 38-48.
- Mandler, G. Response factors in human learning. Psychol. Rev., 1954, 61, 235-244.
- Mandler, G. From association to structure. Psychol. Rev., 1962, 69, 415-427.
- Miller, G. A. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev., 1956, 63, 81-97.
- Miller, G. A. Decision units in the perception of speech. IRE Transactions on Information Theory, 1962, It-8, 81-83.
- Noble, C. E. Meaningfulness and familiarity. Pp. 76-119 in Cofer, C. N. and Musgrave, B. S., eds. Verbal behavior and learning: problems and processes. New York, McGraw-Hill, 1963.
- Pillsbury, W. B. The reading of words: A study in apperception. Amer. J. Psychol., 1897, 8, 315-393.
- Rock, I., & Steinfield, G. Methodological questions in the study of one-trial learning. Science, 1963, 140, 1-2.

Underwood, B. J., & Keppel, G. Coding processes in verbal learning.
J. Verbal Learning and Verbal Behavior, 1963, 1, 250-257.

Underwood, B. J., & Schulz, R. W. Meaningfulness and verbal learning. New York, J. B. Lippincott & Co., 1960.

THE EFFECTS OF MOTOR PRETRAINING ON
CHILDREN'S ASSOCIATION AND RECOGNITION LEARNING*

Harry Levin
Cornell University

Before summarizing several studies concerning stimulus pretraining, I should like to tell you about the setting in which this research problem arose. As many of you know, there is considerable controversy in America about how young children should learn to read. I am emphasizing that the discussion revolves around the word "should" because we have very inadequate information about how children actually do learn to read. Stirred by the continual attention to this problem in the American mass media, a number of us at Cornell University have designed a series of studies to yield information about processes which we believe to be basic to how the child acquires this very complex and obviously useful skill. Since I am talking now in an international setting, I think that it is interesting to note that many educational systems of other nations are reputed to be free of very wide individual differences in the acquisition of reading skill. If these national differences are true, one might look at them for hypotheses about why some children learn to read adequately and others do not. We, of course, first must take into account the capacities of the child. But given this obvious source of difference, one wonders why Italian children, for example, learn to read more easily than American children. A variety of sociological reasons such as the types of children who attend various kinds of schools or the nature of the pedagogical process are apparent. However, for the psychological researcher, a considerably more interesting area of

* Paper delivered at XVIth International Congress of Psychology, Brown, Germany, August, 1960.

speculation and hypothesis concerns the relationships between the nature of the language and the course of learning. For example, certain of the Romance languages possess a high correlation between orthography and pronunciation. English, on the other hand, as many of you who have learned it as a second language know, is beset with many irregularities between orthography and pronunciation which are covered by complex rules or no rules at all. The effects of such seeming anarchy on learning to read is a fascinating area of research, and one that is well worth pursuing.

As is often the case when one approaches an applied problem, such as learning to read, we found that the answers may profitably lie in a number of basic research problems. For example, there is little information about the development of the perception and discrimination of graphic letter-like forms. Some of the research which we started on this problem has been described earlier in these meetings by Dr. Eleanor Gibson. Likewise, since the child brings to his school a well developed spoken language, we need to know substantially more about the development of language than is now available. I have mentioned the problems of regularities and irregularities between writing and speaking and an inductive analysis of the nature of these in English has been undertaken by the linguists on our project. To take a final example, since an important aspect of reading is the process of association learning between a visual stimulus and a verbal response, we felt that here, too, were some fundamental problems in developmental psychology.

The research which I will describe today has to do with conditions which facilitate association learning by children. Specifically, some conditions of stimulus pretraining, or as it is sometimes called, stimulus predifferentiation, were investigated as they relate to subsequent association learning.

Incidentally, I shall prefer the terms "stimulus pretraining" because "predifferentiation" supposes a theoretical explanation for the phenomenon which is not wholly adequate.

Stated simply, the problem we were looking at is the effects of prior writing of words or word-like forms on subsequent associations between these graphic stimuli and their verbal associates. With the exception of an excellent study by Attneave, all of the relevant studies have employed verbal rather than motor pretraining. In light of our practical problem, since children usually learn to read and write simultaneously, writing as a motor pretraining act seemed especially worth attention. I shall describe three studies.

When one watches a young child, say five or six years of age, writing, one is struck by the intensity of concentration and the apparent motor difficulty of the act. The child hunches over the paper, grips his stylus tightly, and produces writings which have only a remote resemblance to usual orthography. The first question we asked was whether the difficulty of the writing act would influence the ways in which the child later made associations to the materials he was writing. Consequently, the pretraining involved tracing of the words under two conditions: the words were either large and easy to trace or quite small so that the child had difficulty in moving his pencil over the tiny forms. The stimulus materials were eight English words which were chosen for their common frequency in written English. Each word had five letters. (It is interesting to note that when we selected materials for the experiment we tried three, four and five letter words and the last gave us the best learning scores. I presume that this is because the large number of letters provide more discriminable cues for the child's learning.)

The eight words were presented on a memory drum, with each list of eight arranged randomly. The procedure was to first let the child see the list run through the memory drum and then to give him the pretraining experience. This experience consisted of tracing each of the words three times. Then the learning procedure took place. The subject was given the verbal associates by the experimenter during two presentations of the list; then the subject made the verbal association on six successive presentations of the list. The learning procedure involved immediate correction in that the experimenter repeated the word after the child. If the child had been correct, the experimenter's verbalization acted as reinforcement; if the child was wrong, it provided correction. Following the learning and test trials, the child was given the printed words and asked to circle the one read to him by the experimenter.

Three groups of children with ten in each group were studied. These children were five years old and had not formally started to learn to read in school. The first group traced the words presented in the large, easy-to-trace format. The second group traced the words when they were small and laborious to go over; the third group had no motor pretraining, but simply looked at the words for a length of time equal to the mean pretraining time of the other two groups.

The results indicated that the pretraining group who traced the words which were small had a slight superiority over the other two groups in association learning and were significantly superior to the other groups in subsequent recognition of the words.

How do we explain these findings? Several years ago in a fine summary article, Dr. Arnault reviewed all of the work on stimulus pretraining which had been done up to 1957. He orders the data in terms

of the nature of the pretraining. For our purposes, the pretraining in the experiment I have just described falls under the rubric of relevant stimulus but irrelevant response. That is, the writing response is not the one which the child will eventually learn, but the stimuli are the ones with which he later will have commerce. The whole area of stimulus pretraining is interesting because it is contrary to what we might expect from the classical work on transfer of training. That is, when the stimulus remains the same but the response changes we expect that there will be negative transfer between the two tasks. However, the majority of the stimulus pretraining studies indicate quite the opposite.

There are various theoretical explanations for this phenomenon. Arnault gives the following four theories: pretraining leads to the acquired distinctiveness of cues, or to the reduction in intralist generalization; or to increased attention to cues; or, finally, it increases the meaningfulness of the material. On intuitive grounds it does not seem that the motor, writing experience we gave the children would increase the meaningfulness of the material. On the other hand, it may very well make the cues more distinctive or increase the subject's attention to relevant cues. It is not possible to choose among these explanations. However, we may ask what pretraining conditions will enhance the distinctiveness of the cues or which will induce greater attention to cues by the subjects. The next experiment was designed to answer this question.

The results of this first experiment led to the hypothesis that pretraining facilitates learning most when the pretraining provides opportunity for the child to monitor his own behavior. By "monitor" we mean the availability of feedback from his own responses in the child's action system. The feedback permits the child to ascertain where he is on the path to the goal and to modify his behavior in terms of the discrepancy

between his position and the goal. Taking this point of view, we would expect that there would be more opportunities to monitor his behavior as a function of the amount of feedback available in the action system. Therefore, more feedback would be likely to occur when the child is an active rather than a passive observer and when more than a single sensory modality is involved in his behavior.

The notion of feedback may underlie the common and interrelated theoretical explanations for the effects of pretraining. For example, with more than one sensory modality involved, the subject may more efficiently differentiate relevant cues in the stimulus materials and with his involvement as an actor rather than an observer, we would expect better attention to cues.

In order to test the hypothesis of the relationship between opportunity for self monitoring and efficacy of pretraining, an experiment was set up in which six pretraining conditions were compared to each other and to a control condition involving no stimulus pretraining.

The procedure was exactly the same as in the previous experiment. However, instead of using English orthography, we developed an artificial orthography. This unfamiliar visual material would control for the variations in exposure to English writing which young children have from the mass media and from their early work in the first grade. These artificial symbols were made by a random combination of the basic upper case strokes in English orthography. We called these letter-like forms graphemes and the combinations of the graphemes into word-like groups "grapheme complexes." Each grapheme was coordinated to a phoneme. Thus, we are dealing with eight English words, each coordinated to an artificial grapheme complex.

Subjects in this study were 84 first-grade children, with equal numbers of boys and girls.

The seven conditions which made up the pretraining conditions were as follows:

1. The child simply looked at the grapheme complexes.
2. The child watched the experimenter trace each grapheme complex.
3. The experimenter held the child's hand and traced the visual pattern with the child.
4. The child himself traced the grapheme complexes.
5. The child copied the grapheme complexes.
6. The child's hand was guided through the patterns which he was prevented from seeing.
7. In the control condition the child and experimenter looked through a picture book for a length of time equal to that taken in the various pretraining experiences.

In an intuitive sense, the first five conditions created an increasingly active involvement of the child and an increase in the number of sensory modalities which he used during the pretraining experience.

The results from this fairly elaborate experiment were disappointing. There were no facilitating effects for any of the pretraining conditions compared to the condition of no pretraining. Likewise, none of the pretraining conditions were better than any of the others. The absence of findings from this experiment may have been due to a number of factors. For one, the subjects received only one pretraining experience per grapheme complex. This may have been too little pretraining to have a facilitating effect on later association learning. Related to this factor is the observation that the eight grapheme complexes may have been

too complicated for differentiation to take place with the limited amount of pretraining experience.

The final experiment I will summarize took into account the two suppositions about the lack of findings in the previous study. The subjects were given three pretraining experiences per grapheme rather than one as was done before. Also, the pretraining was not on the total grapheme complex but on the various parts of this complex. There is substantial evidence that children recognize words in terms of initial letters. The next most efficient cues are the final letters. We predicted, therefore, that the greatest facilitation from pretraining would occur if children had experiences with the first letter of the grapheme complexes, the next best effect would occur when they were pretrained on the last grapheme and the least useful pretraining would be the middle grapheme in the complex. Significant findings on the present study would imply that the previous experiment was inconclusive because the total grapheme complexes were too complicated for differentiation of significant cues to take place with the limited pretraining experiences.

The eight grapheme complexes to be learned and the memory drum presentation of the material were the same as in the previous experiment. The major change was in the nature of the pretraining. Depending on the group to which the child was assigned, he traced, three times each, either the initial, terminal, or the middle grapheme. A control group spent a period of time equal to the average time of the tracing looking through a picture book with the experimenter.

The scores analyzed in this experiment were the total number of correct associations in the six test trials and the total number of correct recognitions. An analysis of variance indicated that there was

a significant effect due to the pretraining condition for both the scores. The level of significance was at about 7 percent level. An inspection of the means indicated that motor pretraining influenced learning in the following way: the highest learning and recognition scores were made by the group that was pretrained on the initial grapheme, next on the terminal graphemes, and least on the middle grapheme. This ordering of means was as predicted.

However, the results in the control condition make the interpretation of these findings difficult. For both scores the training on the initial grapheme led to higher learning and recognition scores than did the control group. However, the control group was somewhat better than either of the other two pretraining conditions.

The results of the last experiment I described indicate that when we know the relevant cues for association, pretraining on the specific cues facilitate learning.

The fact that training on the initial grapheme is only slightly superior to no pretraining at all, which, in turn, is better actually than experience with terminal or middle cues, is difficult to explain. Our best conjecture for this result is that when pretraining calls attention to other than the most relevant cues for subsequent learning, it introduces a confusing factor and creates a negative transfer effect. This generalization probably holds true only for complex stimulus materials.

At any rate, it is in the direction of explicating the relationship between amount of pretraining and the complexity of the materials that we are now directing our research efforts.

In summary, I have reviewed four studies which related children's writing to association learning. The problem grew out of a practical

concern with how children learn to read, and was treated in a framework of stimulus pretraining. The research program is continuing, and, at this point, we are able to say that an effortful motor act yields positive transfer effects to subsequent association learning, as does pretraining on the specific cues which are used by children in making visual-verbal associations.

position more often than not leads to unexamined basic questions of the extensions of "basic" findings to new areas. Partly, the interchange between so-called basic and applied research depends on the way in which the problem is posed. Sometimes, the union depends on a deliberate strategy of analyzing the applied question in terms of fundamental and general problems. Throughout, though, we recognize that the basic-applied distinction is arbitrary and difficult to defend.

The investigation of the processes by which young children learn to read has been an especially fertile area of basic research problems in learning, perception, and language. The present study considers an obvious classroom pedagogical procedure in terms of the laboratory research on stimulus pretraining.

An early aspect of reading involves the attachment of verbal significance to what at the beginning may be nonsense graphic signs. The process is that one of association learning between the orthography and the spoken word. Also, according to common reports of classroom procedures the child at about the age five begins to copy or to reproduce from memory the orthography. Although there have been studies about how the reading and writing processes might relate to each other (e.g., Spalding and Spalding, 1957), the procedures are profitably highlighted. The present study tests the notion that writing may probably

Writing as Pretraining for Association Learning

Harry Levin and John S. Watson and Margaret Feldman

Cornell University

Ithaca College

Our experience has been that research on a presumably applied problem more often than not leads to unanswered basic questions or to the extensions of "basic" findings to new uses. Partly, the interchange between so-called basic and applied research depends on the way in which the problem is posed. Sometimes, the union depends on a deliberate strategy of analysing the applied question to yield more fundamental and general problems. Throughout, though, we recognize that the basic-applied distinction is arbitrary and difficult to defend.

The investigation of the processes by which young children learn to read has been an especially fertile lode of basic research problems in learning, perception, and language. The present study considers an obvious classroom pedagogical procedure in terms of the laboratory research on stimulus pretraining.

An early aspect of reading involves the attachment of verbal utterances to what at the beginning must be nonsense graphic signs. The process is then one of association learning between the orthography and the spoken word. Also, according to common reports of classroom procedures the child at about the same time begins to copy or to reproduce from memory the orthography. Although there have been statements about how the reading and writing processes might relate to each other (e.g., Spalding and Spalding, 1957), the procedures are profitably haphazard. The present study tests the notion that writing may probably

be used as pretraining for subsequent association learning.

According to Arnoult's (1957) fine review of the pretraining literature, motor pretraining, as in writing, has been used only in one study by Attneave (1957). Most of the pretraining studies have used verbal rather than motor pretraining (e.g., Spiker, 1956). Directly relevant to the present study is one by Muehl (1960) who pretrained kindergarten children on a delayed matching task prior to learning a four word vocabulary list. Pretraining on the same words as in the learning task yielded superior results to pretraining on different words or on geometric forms.

The stimulus pretraining area is interesting because it is contrary to what we might expect from the classical work on transfer of training. That is, when the stimulus remains the same but the response changes, we expect that there will be a negative transfer effect between the two tasks. The majority of stimulus pretraining studies, however, indicate quite the opposite.

There are various theoretical explanations for these phenomena. Arnoult (1957) gives the following four theories: pretraining leads to the acquired distinctiveness of cues; or to the reduction of intralist generalization; or to increased attention to cues; or, finally, it increases the meaningfulness of the material. On intuitive grounds, it does not seem that the visuo-motor, writing experience would increase the meaningfulness of the material. On the other hand, it may make the cues more distinctive or induce greater attention to relevant cues.

The critical cues in word recognition by children seems to be first the initial and then the terminal letters in the word. Our best evidence for this generalization comes from the study of intralist

confusions in paired associate learning by children. When two or more words shared an initial grapheme (analogous to alphabetic letter, for our purposes), the words were confused with each other more frequently than words sharing terminal graphemes. Finally, words sharing common medial graphemes or sharing no common elements were least confused (Levin and Watson, 1961).

We reasoned, therefore, that pretraining on initial graphemes would positively influence subsequent paired associate learning more than pretraining on terminal graphemes, and that both of these conditions would yield superior learning to pretraining on medial graphemes. Said another way, if attention could be focussed on the most important cues for discrimination within a list, the pretraining should be most effective.

Method

Subjects. The Ss were 40 first grade children, 20 boys and 20 girls, from the Ithaca Public Schools. The children were randomly assigned to four groups of 10 children each, with sexes equally represented in each group.

Procedure. Since the Ss had had varying experiences with English writing through their work in the early part of the first grade, the mass media and by tuition at home, we used an artificial orthography. Eight English words were chosen according to the following criteria: they are equally familiar to first grade children according to Rinsland's word list (1945); they each have two syllables and five phonemes of which the first and last are different within each word and there were eight different first and eight different last phonemes among the eight words. The words were palace, angel, giant, magic, sparrow, Friday, dishes, and water.

Pretraining. There were three experimental and one control pretraining conditions.

Group I: Traced each of the initial graphemes three times successively. The graphemes were presented in a booklet with the three tracings of a single grapheme per page.

Group II: Traced the middle graphemes. (The three middle positions in each grapheme complex were made up of eight different graphemes).

Group III: Traced the terminal graphemes.

Group IV: Spent the average time for tracing--three minutes--going through a picture book with E.

Learning Task. The grapheme complexes were presented on a memory drum which was calibrated so that a new visual stimulus appeared every eight seconds. The tape contained nine groupings of the eight grapheme complexes. In order to control for serial effects the grapheme complexes were randomized within each of the nine presentations.

Each subject was taken individually from his classroom to a separate experimental room where the child was seated at a small table in front of the memory drum, and given the following instruction: "I have some things for you to do and I think you will enjoy doing them. We are trying to find out how children learn to read. But we know that you have been practicing with regular letters and words so we had to make up some new letters. And with these new letters we made up some new words. We are the only ones who know what these mean and we are going to teach you what they mean." The machine is then turned on. The experimenter continues, "Here are our eight

words. Don't they look strange and funny? That is because they are made up with the new letters."

After the eight grapheme complexes had been shown on the memory drum without any verbalization, the machine was turned off and E said, "Before I tell you what these words mean I have something else I would like to have you do." Depending upon which pretraining group the subject had been assigned to, he was given a booklet and told, "Here are some of the letters we used to make up the words which you will learn a little bit later. I'd like you to take your pencil and trace over them."

Following the various pretraining conditions E said "Now I am going to tell you what these words mean. I will say the word first and you say it after me." The memory drum was turned on and E gave the verbal associates for each grapheme complex as it appeared in the aperture. "Try to remember what each word means." This procedure was followed for two trials. Then E said "Now I want you to try and say the word before I say it. If you are not sure of the word, it is good to guess." Six test trials were given, during which E recorded the child's responses. Immediately following the child's response E gave the correct verbal associate. If the child was right, this acted as a reinforcement; if the child was incorrect, it acted as correction. If toward the end of eight seconds the child did not give a response, he was given the correct associate. Following the six trials, E said, "Now I have one last thing for you to do." E gave the child a recognition test booklet. This booklet contained eight pages on each of which the eight grapheme complexes were arranged in a different order. "Take your pencil and draw a circle around the

word that means 'sparrow'." When a child made a circle regardless of the correctness of choice, E said, "That's fine." "Turn the page", etc. This procedure was followed for all eight words.

Results

The scores analyzed in this experiment were the total number of correct associations in the six test trials and the total number of correct recognitions. These scores were subjected to a 2 x 4 analysis of variance according to sex and training conditions. The means for the two scores are presented in Tables 1 and 2. For both the association learning scores and the recognition scores the F tests for pretraining conditions are significant at about the 7% level ($F = 2.30$ and 2.67 ; $df = 3$ and 32). Since a series of a priori predictions had been made about the relative facilitating effects of the pretraining conditions, t-tests were computed among the conditions.

Table 1. Group Means for Total Correct on the Six Association Trials
(n per cell = 10)

Pretraining Group	I Initial Graphemes	II Middle Graphemes	III Terminal Graphemes	IV Control
Mean number correct	22.5	10.6	13.6	15.9

As can be seen in Tables 1 and 2 the means are in the predicted order; that is, the highest learning scores occur when the children are pretrained on the first grapheme, next on the last grapheme, and learning is poorest for those pretrained on the middle graphemes. For both the association and recognition learning scores, the differences between the means of the group practicing on the first graphemes and the group

practicing on the middle graphemes was significant at the one percent level. In addition, on the recognition scores, the group which pretrained on the initial graphemes also recognized significantly more words correctly than the group practicing on the terminal graphemes ($p < .05$). However, it is important to note that none of the mean scores of any of the pretraining groups differ significantly from that of the control group.

Table 2. Group Means on the Recognition Test (n per cell = 10)

Pretraining group	I Initial Graphemes	II Middle Graphemes	III Terminal Graphemes	IV Control
Mean number recognized correctly	5.6	2.6	3.1	4.5

Discussion

Prior to the present study, the writer carried out a number of experiments with various conditions of writing as pretraining. The results were uncertain. Our best guess about the reasons for the inconclusive findings was that the grapheme complexes given the children to write were long and visually complicated so that the relevant cues for subsequent learning were masked. The current experiment indicates that when the relevant stimulus characteristics for association learning are known, pretraining on the specific cues facilitates learning.

The fact that training on the initial grapheme is only slightly superior to no pretraining at all which in turn is actually better

than experience with terminal and middle cues is difficult to explain. We conjectured, in the light of these findings, that when pretraining calls attention to other than the most relevant cues for subsequent learning, it creates confusion which leads to a negative transfer effect. This generalization probably holds true only for complex stimulus materials.

Summary

This study was designed to test the effects of writing as pretraining for association learning. Forty children, six years old, learned to associate eight English words with an artificial orthographic representation of these words. The Ss were divided into four pretraining groups: One group traced the initial graphemes of the "words"; another, the terminal graphemes; the third group traced the medial graphemes; and a control group had a warm-up period.

The group trained on the initial graphemes had the highest number of correct associates and performed best on a recognition test. Next were the Ss trained on terminal graphemes. The group trained on the medial elements did most poorly. However, the control group performed at a level between the initial and terminal grapheme-trained groups. The results are discussed in terms of the relevance of pretraining to the specific cues used in association learning.

References

- Arnoult, M. D. Stimulus predifferentiation: some generalizations and hypotheses. Psychol. Bull., 1957, 54, 339-350.
- Attneave, F. Transfer of experience with a class-scheme to identification-learning of patterns and shapes. J. exp. Psychol., 1957, 54, 81-88.
- Levin, H. and Watson, J. The learning of variable grapheme-to-grapheme correspondences. Mimeo, 1961, 23 pp.
- Muehl, S. The effects of visual discrimination pretraining on learning to read a vocabulary list in kindergarten children. J. educ. Psychol., 1960, 51, 217-221.
- Rinsland, H. D. A basic vocabulary of elementary school children. New York: Macmillan, 1945.
- Spalding, P. D., and Spalding, W. T. The writing road to reading. New York: Whiteside and Morrow, 1957.
- Spiker, C. C. Experiments with children on the hypothesis of acquired distinctiveness and equivalence of cues. Child Development, 27, 253-261.

time. First the children were told the names of the words, then they were told to write the words, and then they were told to read the words. Finally, they were told to write the words. The question still remains whether or not this is a meaningful measure of the children's learning.

Levin, Watson, and Feldman (see previous paper in this volume) investigated this relationship in terms of writing as a pretraining experience. However, in the present study our interest centers on spontaneous reading and writing since this appears to be a more

This is a revised version of a dissertation submitted to Cornell University for the Ph. D. degree. The complete thesis is available from the Cornell University Library.

THE EFFECT OF INTERPOLATED WRITING

ON PAIRED-ASSOCIATE LEARNING¹

Linda J. Johnson

This study is concerned with the relationship between reading and writing. That the two learning processes are closely associated is apparent in today's schoolroom. Children are taught to read and write at the same time and often children are learning to read and write the same words. Johnny learns to read the word "cat" at the same time that he learns to write it.

Naturally, Johnny needs to learn how to write, but the question seems to be whether or not he should develop this motor skill by practicing on the same materials which are involved in developing quite another skill. One might argue that the child is receiving the same information through two modalities--visual and kinesthetic--and that this dual presentation makes learning more efficient, but is Johnny's written rendition of "cat" actually helping him learn to read the printer's version of the same word?

In some schools an attempt has been made to change this situation. First the children trace over the outline of the word, then they join dots together and form the outline of the word and, finally, they copy the word. The question still remains whether or not this is a meaningful liaison of two learning tasks.

Levin, Watson and Feldman (see previous paper in this report) investigated this relationship in terms of writing as a pretraining experience. However, in the present study our interest centers on simultaneous reading and writing since this appears to be a more

1. An abridged version of a dissertation submitted to Cornell University for the M. S. degree. The complete thesis is available from the Cornell University Library.

appropriate model for what actually occurs in practice. One could not replicate, in an experimental situation, the actual process of learning to read and, therefore, an attempt was made to simulate--as close as possible--the process of learning to read.

It was felt that a paired-associate task would most resemble the situation of learning to read. The child would be shown the visual stimulus and then would be given the appropriate verbal associate. This would be, therefore, a visual-auditory learning task. In likening this experiment to the real-life situation, it was important that the visual stimuli be relatively unknown to the subjects and that the verbal associates, on the other hand, be readily available in the children's vocabularies.

Perhaps it ought to be mentioned at this point that the "writing" involved in this investigation is actually the tracing activity mentioned previously as a first step in learning to write words. It was felt that this type of writing activity would involve the least amount of individual variation in addition to being the most expedient method in an experimental situation. Such being the case, the question asked was whether or not the point at which the tracing activity is introduced in the learning task makes any difference in the subsequent learning. In other words, do children perform differently if tracing occurs early, middle, or late in learning? The experimental design was now a visual-motor activity interpolated at various times in the visual-auditory learning task.

In summary, this investigation attempts to replicate, in an experimental situation, the effect of writing on the process of learning to read.

GENERAL PROCEDURE

I. Sample.

The subjects were 60 1st-grade children, 30 boys and 30 girls, enrolled at South Hill Elementary School in Ithaca, New York. We may assume that the subjects came from middle class homes since this particular public school generally draws its pupils from a middle class neighborhood.

The data were collected during the final 2 weeks of the school year so that all of the children had completed at least 1 year of schooling.

The Ss were randomly assigned to three groups, 20 children in each, with the restriction that the sexes be evenly represented. The nature of these groups is explained below.

II. Materials

The task involved learning eight English words, each of which consisted of two syllables and five phonemes. The words were matched for difficulty according to Rinsland (1945).

It was acknowledged at the outset that the Ss already had varying amounts of experience with the Roman alphabet. Therefore, as in the previous study (Levin, Watson and Feldman), an artificial orthography was used. The unit of the artificial orthography is termed a grapheme and is defined as a letter-like form with an assigned sound which remains constant in all words. Ideally, a one-to-one relationship exists between the phonemes comprising the English words and these graphic representations. However, previous experience indicated that groups of five graphemes were too difficult to learn. Consequently,

in this study, each of the eight words is represented by three graphemes--referred to as grapheme complexes.

ANGEL	2 3 C	MAGIC	π 3 X
DISHES	1 T P	PALACE	4 C 1
FRIDAY	4 9 X	SPARROW	1 4 4
GIANT	8 1 W	WATER	2 W Y

It should be noted that the first and last phonemes are different both within and between words. This being true for all of the words, intra-list confusion due to similar phonemes occurring in similar positions is eliminated. It should also be noted that graphemes represented the first, third and fifth phonemes of each word.

For the learning task, each grapheme complex appeared on a memory drum calibrated to present a different graphic representation every 8 seconds. The complete list of eight grapheme complexes appears 9 times, one instructional trial and eight learning trials. In order to prevent serial learning, the eight grapheme complexes were presented in seven different random orders. Even though the first and second orders appear twice (as the first two and the final two trials of the learning task), the length of the time elapsed was assumed to eliminate any serial effects.

In addition to the list of words on the memory drum, there were sheets of paper with the words in a massed presentation as well as in a distributed presentation. The former type of presentation was used for the interpolated activity and consisted of each word appearing on a single sheet 2 times. The latter type was used for the recognition task and in this case all eight words appeared together on a single sheet.

III. Procedure

A. Learning Task

Each subject came to the experimental room where the task was administered individually by one experimenter. Each of the eight words appeared once and E gave the appropriate verbal associate. S repeated the verbal associate after E, but sometimes had to be encouraged to do so. On all of the subsequent eight learning trials, S was given approximately 8 seconds in which to respond to each of the grapheme complexes with the correct verbal associate. After S made any response, E would immediately give the correct response. If S made no response, E would give the correct response towards the end of the 8-second presentation on the memory drum. In this way, there was immediate reinforcement (both positive and negative) for each of the subject's responses.

B. Interpolated Activity

It was mentioned before that Ss were randomly assigned to three groups. These groups differed only in the point at which the interpolated activity occurred. One group was interrupted after two learning trials, one after four and the third group after six trials. These groups may be labelled "2," "4," and "6," respectively.

The Ss were given sheets of paper which had two representations of one of the grapheme complexes printed on it. Each child was exposed to one sheet for each of the eight words. The task involved tracing four of the words and looking at the other four. The two types of activity were alternated

and always followed the pattern--trace, look, trace, look, trace, look, trace, look. At the start of the interpolated activity, E decided which words S would trace and which words he would look at. This decision involved dividing those words learned prior to interpolation evenly between the tracing and looking activities. In other words, if S had learned four words prior to the interruption, he was asked to trace two of these and look at the other two. In addition, an attempt was made to have each word occur approximately the same amount of times in each of the positions (first, second, third, etc.).

During the interpolated activity, the verbal associates to the materials being traced and looked at were not given by E. S was not encouraged to give them either. Each group of Ss was given the appropriate number of trials in order to total eight.

C. Recognition Task

At the end of the eight trials, the Ss were given a booklet with eight sheets of paper. The eight words appeared in random order on each of the sheets of paper. S circled the word which E read to him. The order of the words to be circled on each page was constant for all Ss.

IV. Scoring

All responses which the Ss made to the various visual stimuli were recorded trial by trial. The scores being used are the number of correct responses and the types of errors.

The rules for responses being scored as correct were as follows:

1. Those responses which were absolutely correct (tense, number, and so on).

2. Those responses which occurred before E gave the verbal associate. Those which occurred concurrently with E's verbalization were not counted as correct.
3. Those responses which were corrected by S before E had a chance to speak. The last response was always the one used for scoring purposes.

RESULTS

I. Effects of Interpolation

These analyses concern the effects attributable to the point at which the interpolation occurred.

A. Immediate Effects

The first analysis concerns the differences in the subjects' scores before and after interpolation. The difference score is obtained by subtracting the score on the trial immediately following the interpolated activity from the trial immediately preceding it.

The difference score eliminates the effects of increased facility with the material since those interrupted after the sixth trial will, in an absolute sense, have learned more than those interrupted after the fourth trial, etc. This score is not concerned with the absolute scores, but rather with the relationship between the trials preceding and following the interpolated activity regardless of the position of occurrence of that activity. An analysis of variance yielded no significant relationships between the difference scores as a function of the point of interpolation.

B. Overall Effects

In addition, an attempt was made to determine the effects of the point of interpolation on the entire course of learning.

In analyzing the differences in overall learning for the three groups, the sum of trials 1+2 was subtracted from the sum of trials 7+8. This score, therefore, indicates the extent of the learning for each of the conditions with initial ability ruled out as a determining factor. An analysis of variance indicates that the three groups do not differ significantly. However, there is a tendency for condition "2" > "4" > "6".

C. Recognition Effects

Aside from the effects of the interpolation on learning, the influence of the point of interpolation on the recognition scores was investigated. An analysis of variance indicates no significant differences among the three groups.

It might be mentioned at this point that there is a correlation between subjects' scores on the learning task and on the recognition task. The correlations are .68 for Condition "2", .69 for Condition "4" and .84 for Condition "6".

D. Summary

In summary then, there are no significant effects attributable to the point of interpolation. The only point which might be relevant in view of the subsequent analyses is the tendency for the Ss in Condition "2" to perform better on the entire learning task than those in Condition "4", and the latter, in turn, do better than those in Condition "6". Although there is a relationship between the learning scores and the recognition scores, this tendency does not hold for the mean recognition scores for each condition.

II. Effects of the Type of Interpolated Activity

In this section, the differential effects of tracing and looking will be discussed.

A. Immediate Effects

Dealing once again with the difference between the trial following and the trial preceding interpolation, the mean difference score for each subject was computed separately for those words which were traced and those which were looked at.

Table 1 indicates the group means for this analysis. An analysis

Table 1. Mean Differences Before and After Interpolation as a Function of Tracing and Looking

Point of Interpolation	Tracing	Looking	Total
"2"	5.35	4.75	5.05
"4"	5.20	4.95	5.08
"6"	5.25	5.05	5.15
Total	5.27	4.92	5.09

of variance by condition and activity was computed for these difference scores and is significant at the .01 level. This analysis indicates that the tracing and looking activities produce differential effects on the subjects' learning. Tracing leads to a decrement in learning immediately following the interpolated activity whereas looking leads to an increment.

Before concluding this section one other analysis must be mentioned and with it the possible implications for our findings. It was apparent that the subjects were performing better on the words which were traced both before as well as after interpolation.

Our present concern is with the fact that the subjects were initially superior on the words which they were subsequently to trace. In other words, the experimenter more often had the subjects trace words which were already known although an attempt had been made to divide the "known" words evenly between the tracing and looking conditions. Therefore, the subjects' scores prior to interpolation were analyzed in terms of the differences between the words which were subsequently traced and those which were subsequently looked at.

A "t" test was computed on the scores prior to interpolation for each condition. The group means are given below:

	<u>Mean of Tracing</u>	<u>Mean of Looking</u>	<u>2-tailed Test</u>
"2"	1.70	11.25	<.05
"4"	4.80	3.75	<.20
"6"	9.05	7.25	<.02

Although the difference is not significant for Condition "4", the significant differences in Conditions "2" and "6" indicate that this initial superiority on the words subsequently traced needs to be taken into consideration.

Due to the fact that the subjects had more correct responses prior to interpolation on the words which they subsequently traced than on those which they subsequently looked at, the possibility arises that the regression phenomenon was operating to cause the losses following tracing and the gains following looking.

In order to check the extent of this regression, if any, a selected subsample for each condition was chosen so that the

mean performance before interpolation was equal for the words which were to be traced and those which were to be looked at. It should be noted that of the 20 subjects in each group, 11 are here included for Condition "2", 14 for Condition "4", and 13 for Condition "6". Therefore, this subsample consists of a fairly large proportion of the original subjects.

It seems obvious that some regression is responsible for the losses following tracing and the gains following looking which were found previously. A comparison of the two samples indicates that the losses are no longer great following tracing and the gains following looking are also not as extreme. However, in all conditions this selected group of subjects consistently performs better immediately following interpolation on those words which were looked at. It is interesting to note that looking produces a gain only for Condition "2" and for this group tracing produces no change. In Condition "4", it is looking which results in no change and tracing produces a loss. Finally, in Condition "6" both activities produce loss, but the loss is more extreme following tracing. Therefore, although the regression phenomenon may be held responsible for some of the differences found between the tracing and looking activities, the superiority of looking as an interpolated activity is still evident.

B. Overall Effects

Besides the immediate effects of the interpolated activity, we are concerned with the overall course of learning as a function of the interpolated activities. The Ss generally have more correct

responses for the words which were traced. The purpose of this part of the analysis is to determine if these differences are, in fact, significant.

As a first step, the total number of correct responses for both activities were compiled and a "t" test for comparison was done separately for each condition. The results may be summarized as follows:

	<u>Mean of Tracing</u>	<u>Mean of Looking</u>	<u>2-tailed Test</u>
"2"	15.15	12.90	<.05
"4"	13.30	11.70	<.20
"6"	13.70	11.65	<.02

The results are significant for Conditions "2" and "6" and thus indicate that the subjects in these two groups do better overall on the words traced than on the words looked at. Eliminating those trials occurring prior to interpolation, "t" tests were again computed for each condition separately and this time included only those correct responses occurring after the interpolated activity. The results are as follows:

	<u>Mean of Tracing</u>	<u>Mean of Looking</u>	<u>2-tailed Test</u>
"2"	13.45	11.65	<.05
"4"	8.50	7.95	<.50
"6"	4.60	4.45	<.60

With this correction, the results are significant for Condition "2" only. This again indicates that differences seem to be more extreme for those subjects experiencing early interpolation.

An additional analysis was done in this general area of the extent of overall learning. Once again an individual was given a score for Trials 7+8 minus Trials 1+2. However, this time the scores were computed separately for those words traced and those words looked at. When the final learning was corrected

for initial ability it was found that there is no significant difference between the words subjected to the tracing activity and those subjected to the looking activity.

In summary, the tracing activity does not produce significant differences in overall learning when correction is made for initial ability. Any differences in favor of the facilitating effect of tracing on later learning can be attributed to the fact that during the interpolated activity the subjects more often traced words which they already knew.

C. Recognition Effects

In regard to the effect of the interpolated activity on recognition, a χ^2 analysis was done to determine if there is a difference due to tracing and looking on each of the conditions. The analysis showed no significant differences.

D. Summary

In concluding this section, it appears that the type of interpolated activity definitely affects the course of the subjects' learning immediately following interpolation. The decreasing effect of tracing and the increasing effect of looking, the result of an analysis of variance of the difference scores, is interpreted to be due, in part, to the regression phenomenon. This interpretation was arrived at due to the inequality of the subjects' scores, prior to interpolation, on the words to be traced and the words to be looked at. When a group of subjects performing equally well on the words subsequently traced and those subsequently looked at was selected,

the extreme losses following tracing and the extreme gains following looking were no longer evident. However, in all conditions the looking activity always resulted in more superior learning than the tracing activity. In addition, this selected sample seems to indicate that interpolation is more efficient when it occurs early. Condition "2" is the only group in which there was an actual gain following looking and no change, rather than a loss in the other two conditions, following tracing.

When correction is made for the initial superiority on the words to be traced, there are no differences in overall learning between the words traced and the words looked at.

As far as recognition is concerned, there are no significant differences between the two interpolated activities either by condition or by word.

The Learning of Variable Grapheme-to-Phoneme Correspondences

Harry Levin and John Watson

Cornell University

Many children find that learning to read English is hard. Since substantial concern has been paid to those children who actually have so much difficulty that they are classified as retarded or problem readers, the focus of research has been on the individual differences which predict reading disability. Intelligence, sex, lateral dominance, and perceptual acuity are representative variables and their relations to reading are adequately reviewed by Robinson (1946).

Recently, another tack has been suggested by linguists and reading specialists who have compared the acquisition of reading skill in languages other than English. The comparisons are informal and speculative. Nevertheless, languages may be ordered in terms of the regularity of their phoneme-grapheme correspondences so that, according to Bloomfield (1942), Spanish, Bohemian, and Finnish have almost perfect alphabetic systems; Italian, Dutch, and German are relatively imperfect, and modern Greek, French, and English are extremely imperfect and arbitrary. The important point, moreover, is that the incidence of reading disabilities varies inversely with the regularity of the alphabetic systems, but this generalization is little more than a guess.

As fascinating as the historical reasons for imperfections in the writing-to-sound systems may be, they are not the concern of this paper. The one obvious point, though, is that writing systems are Johnny-come-latelies in comparison to spoken systems and the changes in language and their alphabets have often been asynchronous. For our purposes, these varied regularities provide many interesting research problems and suggest avenues of handling reading skill problems via the presentation of materials as well as concern with the characteristics of the neophyte reader.

English is a prime offender in irregularities. Soffiatti (1955) points out after Fries that spoken English has 35 phonemes or basic sound contrast units. Yet, ...not a single one of the 35 phonemes of our language is represented in writing with only one spelling. As a matter of fact, some have up to thirteen spellings (p. 72). Several examples may emphasize this point: the phoneme /i/ may be spelled ee, e, es, se, oe, eo, ei, ie, i, ey, ay. Or again /s/ as in ship has the following spellings: sh, ce, ch, ci, pah, s, sch, sci, se, si, ss, sei, ti.

Some of these variations are regular, that is, predictable in certain forms. In the second of the above examples, sh is pronounced in only one way. Other spelling forms, although they are irregular, may be handled in a large number of instances by more or less complicated rules. And, unfortunately, some forms have to be learned by rote. These are the circumstances which make reading English difficult and the difficulty itself creates large individual variations in reading skill.

What implications for the reading of English are posed by these linguistic facts? Bloomfield (1942) and Soffiatti (1955) make these unequivocal recommendations. Train the child first on the regular forms and then introduce the irregularities.

Our first reading material must show each letter in only one phonetic value; thus, if we have words with g in the value that it has in get, gets, gun, our first material must not contain words like gem....

The second stage takes up regular spellings in which double consonants and other digraphs appear in consistent uses, e.g., ll as in well, th as in thin....

The third stage takes up words whose spellings may be called semi-irregular, for example, the type of line, shine, mile, while.... (quoted by Soffiatti, p. 79).

Although these recommendations may indeed follow from the nature of English phonology and writing, they may be questioned from the point of view of the learner. This incompatibility is not uncommon. Oftentimes, the logic of the

subject matter dictates sequences of materials to be learned, yet these sequences may be less than optimal for efficient acquisition or transfer. Especially in terms of transfer, the order of presentation dictated by the logic of the subject matter may not be the best for other content which follows.

Stated in its simplest form, Bloomfield is suggesting that the child learn at first to make only one verbal response to a given grapheme. After this response is overlearned, the same graphic representation may acquire other associated sounds. A number of psychological principles and findings may be brought to bear on this question. First consider the acquisition process itself; that is, forming the association between the graphic representations and the verbal responses. The results of much research on paired associate learning may be brought to bear. If the contrasts among the elements in the list are large, discrimination among the elements of the list is facilitated and, consequently, the lists will be learned more rapidly (Rothkopf, 1958). This suggests that if differences among the graphic elements or among the responses to these graphemes are available these variations will provide a basis for discrimination among the elements and consequently should lead to faster acquisition. The second part of this paper will present evidence on the nature of the confusions occurring as a function of intralist similarities.

Besides acquisition criteria, we may consider the conditions facilitating transfer which, in the practical reading situation, is an equally or even more important criterion. Two aspects of transfer are relevant. One is the recognition of the previously learned graphic elements in new contexts. The best evidence that we may extrapolate to the current situation is that training with maximal perceptual contrast will yield the largest positive transfer. Baker and Osgood (1954) point out that transfer is greatest when the criterion task is presented in steps of increasing difficulty. On the face of it, these findings seem contrary to our hypothesis. However, these researchers found that training which is easy leads to deterioration in subsequent list performance, perhaps

because of attitudinal changes in the subjects. These findings provide a ready bridge between the two types of transfer we are considering.

In order to understand the second aspect of transfer, one may think of reading as a decoding process. As Bloomfield and other linguists have pointed out, the early aspects of learning may be thought of as the translation of the graphic code into its verbal equivalent. The job for the beginning reader, then, is to decode the orthography into combinations of sounds which the child will recognize. This assumes that the code stands for language which the child has available in his repertoire. When he produces sounds from the code, he will recognize them as they approach words or other verbal elements he has previously heard or used. Since the child obviously approaches reading with a substantial verbal vocabulary, it is not unlikely that the decoding process will yield familiar elements to him. This approach to reading holds in abeyance the problems of reading skill such as speed, blending of elements, etc.

Reading as a decoding process presents a performance problem for the child. Suppose he encounters elements in novel combinations. Suppose, further, that one or more of the elements has more than a single verbal associate. Assume, also, that the child sounds out the elements, but they do not produce an utterance with which he is familiar. The training problem is to get him to resound the elements, giving those with variable possibilities other values to see whether the second combination will yield a word that he has in his repertoire. This process is akin to problem solving. In problem solving the child may try a solution which yields an answer or his solution may be a failure. The attack on the problem would need another solution. In general, what we want to train the child for is an approach of variability and this might be thought of as the transfer of a general principle or approach.

How do we train children for variability? We may think of the process as one in which the child tries the first habit in a hierarchy which does not work and then tries a second habit and so on until a solution is successful.

We might suppose that the closer the two habits are in the hierarchy, the more likely will a succeeding one be substituted for an unsuccessful response. Returning to Bloomfield's suggestion of overlearning on a single form, we might predict that if the first does not work, the distance between habits in such a hierarchy will make it less likely for the second which is much weaker to be applied. This also suggests that various appropriate habits be learned both to more or less equal strength and simultaneously in time.

An experiment by one of the authors is relevant (Ackerman and Levin, 1958). In order to create a set toward variability in children, the subjects were in one case taught two solutions to one problem and in another, a single solution to the problems. Those taught alternative solutions did better on similar and widely different transfer problems. Other studies at least indirectly support this conclusion (Dashiell, 1924; Crafts, 1927; Callantine and Warren, 1935; Maltzman, Bogartz, and Breger, 1958). We are extrapolating such findings to suggest that in order to maximize transfer of such an attitude the various verbal responses to a given grapheme should be taught simultaneously.

To this point we have painted Bloomfield's and our hypotheses about acquisition and transfer as being diametrically opposed. Obviously, this is not the case although our prediction is that some degree of variability is more useful than Bloomfield's categorical training in single stimuli and responses. The large numbers of complicated associations which exist in English must be taken into consideration in terms of children's capacities to handle degrees of complexity. Should some variability as introduced in the studies to be discussed yield results in line with the hypotheses, we still would be reluctant to say that the total range of variation should be imposed on the beginning reader at the same time. In fact, Morrisett and Hovland (1959) found that a small number of different problems was better than a large number which in turn yielded poorer transfer than training in single types of problems.

In spite of the various complexities in the above discussion, the essential elements of the issue may be abstracted and operationalized without great difficulty. The subjects in the study to be described were in one case given lists of words which had a constant grapheme-to-phoneme correspondence, followed by lists in which this relationship was variable. Another group started with a variable list followed by another variable list. Both groups finally learned a common variable transfer list.²

Method

Subjects. The subjects were 40 third-grade children, 20 boys and 20 girls.

Materials. The lists which the two groups learned as well as the transfer list which was common to the two groups is given in Table 1.

Table 1. Lists Used for Various Conditions

Constant	Variable	Variable	Variable	Transfer List
Hit	Rate	Hit	Rate	Fog
His	Rose	His	Rose	Fan
Sit	Late	Late	Sit	Dan
Limb	Comb	Comb	Limb	Dog

These words are the English equivalent of graphic representations which were rendered in the artificial orthography. It will be noted that a total of eight words made up the various training lists. Each of the words is a monosyllable of the form, consonant-vowel-consonant. One grapheme is related to each phoneme. The variations in the lists occurred in the medial, vowel position. In the constant list, a single grapheme stood for the single medial vowel /i/. The variable lists contained either two or three different vowel

2. This epitomizes our strategy in dealing with complex, applied problems. First, the theoretically and empirically basic elements of the problem are abstracted and investigated under simple, controlled conditions. Should the results merit it, the next steps may involve testing the propositions in the more complicated, real situations.

sounds. However, all of the vowels were represented by only one graphic symbol. The transfer list included different vowels and consonants from those in the acquisition lists. Also, in this list, the medial vowel had only one graphic representation for the two sounds represented.

Further, the lists bear various similarities to each other. Consider, for example, the first constant and variable lists. These two lists have two common words and two words that are different. The same is true for the two variable lists that come second in the learning sequence. This arrangement permits the analysis of errors on the common words as well as the uncommon words, equating the amount of previous learning. Also, there are common words across lists. The analysis of learning of these words is confounded by the fact that they may have appeared on the first or the second list for the various groups. For example, sit and limb appear both in the first constant and in the second variable lists.

Procedure. Twenty Ss were randomly assigned to each of the two learning conditions, with the restriction that sexes be represented equally in each of the groups. The first group will be called the constant-variable group (C-V), and the other group is a variable-variable one (V-V). The detailed instructions to the subjects were as follows:

Introduction: When S arrives at the testing session, the experimenter (E) introduces himself and says, "I have some things for you to do and I think you will enjoy doing them."

Acquisition Learning Session: Step #1: S is seated and E says, "Before we begin, I would like to tell you why you have been asked to come here today. The reason is this. I want to find the best way to teach children how to read, and you can help me find the answer to this question."

"You see these funny looking letters? (E shows S a sheet of paper containing the 12 letters of the artificial orthography). I made up these brand new letters because you have already been working with real letters in school. Now with these new letters I made up some brand new words that nobody has ever seen

before. And I am going to give you a chance to learn four of these new words right now. O.K.?

"I will put the word up here (E points to card platform on a table directly in front of S, E being seated at the table to the right side of S), and I will tell you what the word means. You repeat the word after me so that I will know you heard me. Look at each word carefully and try to remember it. O.K., here we go. Here is the first word - This is _____. (After S has repeated the associate, he is given about five seconds to look at the tri-graph). O.K., here is the second word, this is _____O.K., here is the third word, this is _____O. K., here is the fourth word, this is _____.

"Those were the four words. Now we are going to go through the words again, but this time I want you to try to say the word before I say it. I'll say it after you. If you are not sure what the word is, you can guess. (E shuffles the four cards into prearranged order and then places one on the platform.) What do you think this word is? (Following S's response E will say the correct associate and move on to the next card. If S does not respond within 10 sec., E will say the correct associate and then move on to the next card.) What do you think this word is?..... (This procedure repeated through the fourth card.)

"Now we will go through the words again. (E shuffles the four cards into prearranged order and then places one on the platform, following S's response or following 10th sec. - whichever comes first - E says the correct verbal associate and places the next board on the platform...and so on through all four cards.)

(This procedure is continued until S gets all four associations correct on each of three successive trials or until the fifteenth trial--whichever comes first.)

Step #2: "Well you learned those words very well, very well indeed. Now I'm going to give you a chance to learn four more words. We will do this the same way as last time. First, I'll tell you what the new words are and you repeat

them after me, and then you will have a chance to say the word before I do. O.K., here we go then. This word is _____. "... (The same procedure is followed in this step as was the case in step #1.)

The criterion of learning was the same as on the first list.

Between the trials on the two lists and the start of the transfer list, the E and S looked at a picture book together and conversed for three minutes. The instructions for the transfer learning session were as follows:

Transfer Learning Session: Now I will give you a chance to learn four more words. We will do this the same way as the other two times. O.K., here we go then, this word is _____. "... (The same procedure was as for steps #1 and #2.)

Each of the words were printed in one-inch letters on 3 x 5 white cards. E recorded the actual associations given to each graphic stimulus.

Results

Two learning criteria were used. The first is the number of trials to reach the criterion of three successive correct associations to the list of four words, and the second is the number of errors in reaching the criterion. One word about the way the trials-to-criterion were scored. The trials score for a child was taken as the first of the three trials on which he made the correct associations; that is, if a child was correct on trials 11, 12, and 13, his score was "11" for the number of trials to criterion. If S did not reach criterion, he was assigned a score of "15."

The first comparison on these two criteria is on the first list under each condition; i.e., the C list for the C-V group and the V_1 list for the V-V group. It can be seen in Table 2 that the two groups differ very little on the number of trials to reach criterion on this first list. There is a slight, not significant, tendency for the variable list to be learned with fewer errors ($t < 1.00$).

Table 2. Mean Trials to Criterion and Errors

	C-V Group		V-V Group		t Tests	
	Trials to Criterion	Errors	Trials to Criterion	Errors	Trials	Errors
C List	12.80	20.15	V ₁ List	11.60	N.S.	N.S.
V List	10.65	15.45	V ₂ List	7.90	2.05 p<.05	2.67 p<.02
Transfer List	11.80	15.20	Transfer List	12.05	N.S.	N.S.

The most interesting comparison is the second list for each of the two groups-- the first variable list for the C-V group and the second variable list for the V-V group. It can be seen that V-V group learned more efficiently according to both criteria. The mean trials to criterion for the C-V group was 10.65 compared to 7.90 for the other group ($t = 2.05$, $p < .05$). The differences between the mean error scores shows an even wider variation in favor of the V-V group ($t = 2.67$, $p < .02$). Before taking up the results on the transfer list some further analyses of the performance on the two acquisition lists help to clarify these differences.

Table 3. Mean Numbers of Errors

	C-V	Common Words		t tests	
		V-V			
List 1.	14.20	9.75	2.12	$p < .05$	(Hit, His)
List 2.	10.35	6.35	2.17	$p < .05$	(Rate, Rose)
Exchange Words					
List 1	5.95	(Limb, Sit)	7.70	N.S.	(Comb, Late)
List 2	5.10	(Comb, Late)	2.65	2.60	$p < .02$ (Limb, Sit)

In Table 3 the mean errors are broken down according to the group, the list, and whether the words were common to the two lists or were exchange words. It can be seen that on the first list, the C-V group and the V-V group shared the

words, hit and his. The mean errors was smaller for the V-V group ($t = 2.12$, $p < .05$). The same is true on the two common words in the second lists, rate and rose ($t = 2.17$, $p < .05$). The overall number of errors on List 2 compared to List 1 decreases for each of the groups, which is most likely a learning-to-learn phenomenon.

The other words which were exchanged among the various lists present a more complicated picture. For the C-V group the words limb and sit received fewer errors than the words comb and late on the V-V group, but this difference is not statistically significant. Also, limb and sit which appeared on List 2 for the V-V group have the smallest overall errors. From these results, one might infer that in spite of context and order of the lists, the words, limb and sit were the easiest to learn. The fact that these words occur on the second list for the V-V group contribute to the overall small number of errors for that group. However, the fact that the V-V group on the second list performed better than the C-V group on its second list on the words that were common to both, makes the case stronger that the variable-variable condition results in eventual superior learning.

The transfer list included the words, fog, fan, dan, and dog. The two medial vowel sounds were represented by a single grapheme so that this was a variable list. All of the phonemes and graphemes were new in relation to the previous lists. It can be seen in Table 2 that the two groups did not differ in their rate of acquisition of this transfer list, using either of the two criteria. Several reasons are suggested for the absence of a transfer effect. The list itself because of the internal relationships among the words may have had peculiar effects. Or, the previous acquisition training may have been insufficient to create a "set toward variability." In order to understand the nature of the performance on the transfer list, an extensive analysis of the errors made to that list was carried out. The subsequent parts of this report will deal with the results of the error analysis which, in turn, gave rise to new and somewhat unexpected hypotheses.

Error Analysis

In learning the lists, children made three kinds of errors: 1) omissions, 2) intrusions in which the child gave words which were not on the list; and 3) confusions which was the most common type of error and involved giving a verbal response which was incorrect for that particular visual stimulus but was another word on the list.

Confusion Errors

As can be seen in Table 4, the transfer list lends itself to a symmetrical analysis of confusions. To the stimulus word, dan, confusion error responses would be the other three words on the list, dog, fan, or fog. Likewise, the same types of confusions can be analyzed for the remaining words on the list.

Table 4. Analysis of Confusion Errors

Stimuli	Class of Confusion Errors			Totals
	I	II	IV	
Dan	75 <u>dog</u>	30 <u>fan</u>	14 <u>fog</u>	119
Fan	91 <u>fog</u>	38 <u>dan</u>	24 <u>dog</u>	153
Fog	106 <u>fan</u>	40 <u>dog</u>	16 <u>dan</u>	162
Dog	84 <u>dan</u>	19 <u>fog</u>	14 <u>fan</u>	117
Totals	356	127	68	555
Percent	65	23	12	

It will be noted that the words have a special phonetic and graphic relationship to each other. The word dog shares an initial grapheme and phoneme with dan; fan shares a terminal grapheme and phoneme with dan and fog shares no phonemes or graphemes with dan. We will call an error which substitutes a word with a

common initial sound a Class I confusion; the sharing of a terminal sound, a Class II confusion; and no sharing a Class IV confusion. (The reason for the omission of Class III will be taken up later.)

The 40 children who learned the transfer list made a total of 555 confusion errors. Of these, sixty-five percent were in Class I, 23 percent in Class II, and 12 percent in Class IV. The hypothesis that the errors would be in the order, Class I > Class II > Class IV, was tested in three steps by using the sign test for large samples. Class I exceeded Class II and Class IV at probability levels well beyond .01. Class II had more errors than Class III, with the probability of .02. In other words, the major source of intra-list confusion errors appears to be the fact that two or more words share both the first graphic symbol and its associated phoneme. Next in magnitude of error is the sharing of terminal sounds and graphs, and when words within the list have no common initial or terminal graphemes and sounds, they are least likely to be confused.

The first two lists learned by these groups may also be analyzed for confusion errors in the same way. These lists, however, are not symmetrical. That is, the constant list, for example, has possibilities for two instances of Class II errors and eight possibilities for the Class IV errors. On each of the four acquisition lists in this study, the Class I errors was greater than the Class II which, in turn, was greater than Class IV.

Additional supporting evidence about the major confusions in learning words of this type comes from a second study carried out with another group of subjects³. Subjects for this study were the same age as in the original study and also included 40 children equally divided between boys and girls.

3. We wish to thank Janet Feldtmose, Avrama Gingold, and Jane Hart for collecting and analysing these data.

Table 5. Confusion Errors - Constant Training List

Stimuli	Class			
	I	II	III	IV
bat(1,2,4	71(back)	39(cat)	13(nap)	1(sad) 11(had)
had(2,4)		41(sad)	18(back)	15(nap) 14(bat) 8(cat)
back(1,3,4)	63(bat)		30(cat)	14(nap) 14(had) 9(sad)
sad(2,4)		19(had)	15(nap)	15(back) 8(cat) 5(bat)
nap(4)			14(had)	11(cat) 8(bat) 6(sad) 4(back)
cat(2,3,4)		36(bat)	24(back)	
$\bar{X}$ =	67.00	33.75	27.00	11.50
$\bar{X}$ percent=	48.11	24.24	19.39	8.26

Table 6. Confusion Errors - Variable Training List

Stimuli	Class			
	I	II	III	IV
had(1,2,4)	100(home)	47(mud)	54(nap)	4(but) 2(soap) 3(had)
soap(2,4)		9(nap)	9(but)	5(mud) 0(home)
home(1,3,4)	42(had)		32(mud)	38(nap) 11(but) 2(soap)
nap(2,4)		12(soap)	23(mud)	18(home) 15(had) 13(but)
but(4)			22(home)	19(mud) 12(soap) 8(had) 7(nap)
mud(2,3,4)		35(had)	48(home)	53(nap) 3(soap) 3(but)
$\bar{X}$ =	71.00	25.75	40.00	14.73 151.48
$\bar{X}$ percent=	66.87	17.00	26.41	9.72

The task given to the children was to learn a list of six words to which the possibilities of four types of confusions may result. Besides the three classes of errors, another source of confusion exists in the lists. A pair of words in each list (cat-back and mud-home) are related to each other in that the initial grapheme-phoneme of one word is terminal in the other. Confusions between these words are labelled Class III errors which are predicted to fall in value between Classes II and IV. The first list given to half of these Ss was a constant list in terms of our previous definition of the relation between a single grapheme and a single phoneme. In this case the order of errors was highest for Class I (48 percent) and next for the other three classes of errors in the predicted descending order.

A variable training list, also of six words, which was given to the other half of the Ss shows somewhat less clear results. The most errors (47 percent) were in Class I and the fewest errors (10 percent) were in Class IV. There was, however, a reversal between Classes II and III confusion errors in that there were more Class III than Class II errors.

Table 7. Confusion Errors for Four Lists (Mean percents)

		Class of Errors			N*
		I	II	IV	
C-V	List 1	56.3	33.1	10.6	267
	List 2	44.9	37.6	17.5	231
V-V	List 1	58.5	23.8	17.7	270
	List 2	77.4	11.6	11.0	103

* Total number of confusions.

In summary, we may say that for children the greatest source of confusions in learning groups of words is the case when two or more words share an initial grapheme-phoneme. The next major source of error is the sharing of terminal

phonemes; the next is the commonness of the initial unit of one and the final unit of another word. The least source of confusion exists among words that share neither initial nor terminal graphemes-phonemes.

In light of these findings we may reconsider the acquisition results of the original study and explore the differences between the C-V and V-V groups on the basis of the confusions on the second lists on which the two groups differed markedly. The lists present an equal number of confusion possibilities of Classes I, II, and IV. Actually, on the second list the C-V group had an average of 11.5 confusions and the V-V group only 5.15 confusions. The number of confusions on the first list for each of the two groups was almost equal as was the number of confusions on the transfer list (see Table 7).

Intrusion Errors

The frequency of intrusion errors is much lower than confusions, so that the following discussion is based on relatively few instances and must be considered speculative. Our analysis of the possibilities of intrusion was based on the possible combinations among the phonemes represented in the four words in the lists. For example, the list for the C-V group which included hit, limb, his, and sit has a number of permutations possible. Each phoneme was weighted as a function of its frequency in the list and its position in the word. For example, the phoneme /h/ and its grapheme appears twice in an initial position so that this would be the most probable initial phoneme for the intrusions. Likewise, the most likely terminal phoneme would be /t/ since it appears twice. Also, the phonemes /m/ and /s/ appear once in terminal positions. The possible combinations of these phonemes, then, in terms of the probabilities of occurrence of the phonemes in the list and their combinational possibilities yields the word him as the most likely intrusion. The next most likely intrusion would be lit and other possible intrusions are liz, sim, siz, etc. Overriding these combinational possibilities, some of which yield nonsense words, was the expectation that the

combinations which yield meaningful English words would be intruded more frequently than nonsense combinations. In the lists which we are considering, the word him was intruded a total of 32 times by seven subjects; it was given 12 times to the stimulus hit, 14 times to the stimulus his, and 6 times to the stimulus sit. The interpretation of the intrusion response him to this list is confounded because of the semantic relationship between his and him. All other intrusions were given by only one or two subjects. Therefore, the word sat was given to sit four times, and the word leaf to limb seven times. The most likely explanation for these intrusions is semantic generalization from the stimulus words.

To the list, rate, comb, rose, and late, the most highly predicted intrusion was the word, rote. This was given by two subjects a total of six times: four to the stimulus rate and two to the stimulus rose.

On the first of the V-V lists--hit, comb, his, late--the predicted most common intrusion is again him but it was not used by any subject. However, the intrusions were so infrequent in this list, only 19 instances, that one cannot draw any conclusions about intrusions.

Intrusions are particularly interesting on the second list of the V-V group, on which there were a total of 45. The most frequent predicted intrusion was the nonsense word rit, and it actually occurred twice by one subject, even though it is not an English word. The most frequent intrusion in this list was the word sat given by 4 subjects 9 times to the stimulus sit. This is semantically predicted intrusion.

The total number of intrusions on the four lists learned by the two experimental groups presents some interesting comparisons. On the first list of the C-V group there are 66 intrusions, compared to only 19 intrusions on the first list of the V-V group. However, on the second list which presents the major contrast between the groups the rates of intrusions vary markedly. The C-V group on its variable list has 18 intrusions whereas the variable list of the

V-V group has 45 intrusions. The group of four intrusion totals differs markedly from chance (Chi square = 30.78, $p < .001$).

It will be recalled that one of the consequences of variable training was predicted to be a general set toward variability. The results from the earlier Ackerman and Levin (1958) study indicated that those Ss who had variable solution training to problems presented more solutions to new problems as compared to simply omitting solutions. If we make a rough analogy to the present study, we may say that original variability training leads to at least some attempts to answer with a response to stimuli rather than omitting the response, even though the response may be incorrect. Likewise, the two groups differed in the number of omissions on this list (C-V = 59; V-V = 32).

To summarize the results, variable training on the list appears to enhance the learning of subsequent variable grapheme-phoneme relationships. However, this variability does not transfer to a third variable list. Also, the nature of errors indicates that errors of confusion within the same list tend to be determined by the common grapheme-phoneme combinations existing within the list. The reasons for these results are quite complex and will be discussed in the subsequent section.

Discussion

Based on several lines of reasoning, the prediction was made in this study that lists which contain words having variable grapheme-to-phoneme correspondences would be easier to learn than lists in which the correspondences were constant; and, secondly, that a set toward variability would be acquired in the process of such original learning. In the first lists which compared a constant versus a variable arrangement there was a slight, though not significant, tendency for the variable list to be acquired with fewer errors. The sharpest findings occur on the second list which contrasted two variable lists for the two experimental groups. In this case, the group whose variable list was preceded

by another variable list performed significantly better than the group whose second list was preceded by a constant list. What are the reasons for this obtained difference? The possibility that the lists were intrinsically different in difficulty was not supported because even on the two words that were common to both lists the variable-variable group performed better. The most tenable explanation is that the subjects who recognize the grapheme-phoneme variations in the first list built an expectance for such variations which facilitated the second learning. In other words, we are assuming that the set for variability was engendered by the first list and transferred as a general set to learning the second list. However, by this reasoning one would expect the V-V group to perform better on the transfer list. There are no clear reasons why the groups did not differ on this third list except that the list itself may have been so arranged that the variety of confusion errors possible mitigated the transfer effects. Also, by that time the C-V group has had variable training and experience with as many graphemes and phonemes as the V-V group.

The "set toward variability" explanation is further supported by the larger number of intrusions in the second list of the V-V group and the fewer number of omissions by this group. In a sense, this group was willing to try a new response expecting that some variation of response may lead to a correct one.

The reasoning about the easier learning in the variable list was based on the fact that fewer intralist confusions would occur when there were variations within the lists. Said another way, a list which has bases for discriminating among the words in the list should be easier to learn than a list which does not have such possibilities for discrimination. In this case, the variable responses should provide some basis for discrimination. This hypothesis may be tested by comparing the constant and variable lists, ignoring previous experiences. As was pointed out earlier, there was a tendency for the variable list to be learned with somewhat fewer errors, although the difference is not significant.

When the variability of the list is held constant as in the second list under both conditions and the previous experience is varied then the variable list does seem to make a difference.

The sources of confusion errors are both striking and consistent across various lists and replicable through several studies. From the information at hand, however, we cannot tell whether these confusions are due to similarities in the verbal associates or to the similarities in the graphic stimuli. This leaves open the question of whether the confusion is one in decoding the orthography or whether it is a memory confusion among the various verbal responses. The fact that the initial element of both the stimulus and the response is the major source of confusion points out that the child appears to use the initial element in the word as the major cue for the recognition of the word. Following this in importance is the final element in the word. There have been a number of early studies indicating that the child bases his recognition of words on these positions. A further bit of evidence on this topic was gained by the writers in a previous study on pretraining. In that study the subjects wrote as pretraining either the first or the terminal or the medial element of a word. They then learned to make English verbal associations to these nonsense graphemes. It appeared that training on the initial element was most effective, pretraining on the terminal element next and pretraining on the medial element least effective.

The most likely explanation to us of what happens to these confusions is that the child is able to decode the initial grapheme which triggers an incorrect association. That is, to the initial sound /h/ as in his the child has several associations, is able to decode the h i and has the memory confusion between his and hit. For those children who are unable to decode the initial grapheme the most likely next position in the word at which he makes an attempt appears to be the terminal grapheme. All of this is a way of saying that we are not able to separate the decoding and the memory confusion aspects of these

confusion errors and it may very well be that this is not a meaningful separation.

The confusion analysis does indicate the major part that phonological confusions play in the child's learning a list of words. Surprisingly, in spite of the tremendous attention that has been paid to intra- and inter-list phenomena in paired associate learning, little attention has been given to this particular sound source of confusion. A much more common focus of interest has been on semantic confusions. It is our hypothesis that phonological confusions are more important for the younger child and semantic confusions are more important for the older child. And there is some evidence to indicate that young children do not make semantic generalizations in conditioning as do older children and adults. In line with these speculations, a study is being planned in which the phonological source of confusion is compared to the semantic source with children of various ages.

The confusion analysis also pointed to a next step in the training for variability. In the study just reported the variations between the graphemes and phonemes occurred on the medial vowel. This has obviously turned out to be the position in the word which leads to the smallest number of confusions. Therefore, grapheme-to-phoneme irregularities occurring on either the first or the terminal graphemes and phonemes should provide more potent effects than did the medial position variations used in this study. The next study being planned makes such variations on the initial positions in words.

Summary

Based on speculations about how children might be taught to read, the hypothesis was tested that children more easily acquire and transfer grapheme-to phoneme correspondences which are variable than those that are constant. That is, when a given graphic symbol stands for two or more sounds, it should be learned and transferred more readily when the variations are learned together than when they are learned separately.

Two groups of third grade children learned lists of words in this order: constant followed by a variable list compared to variable followed by a variable list. Both groups then learned a common transfer list of words. The V-V group learned its second list with less trials and errors than the C-V group. They did not differ on the transfer list.

A subsequent analysis of the confusion errors on the various lists indicated that words which share an initial grapheme and phoneme are most confused with each other; sharing terminal elements are next as a source of confusion; and words with no common elements are least confused.

References

- Ackerman, W. I., and Levin, H. Effects of training in alternative solutions on subsequent problem solving. J. educ. Psychol., 1958, 49, 239-244.
- Baker, R. A., and Osgood, S. W. Discrimination transfer along a pitch continuum. J. exp. Psychol., 1954, 48, 241-246.
- Bloomfield, L. Linguistics and reading. Elem. Eng. Rev., 1942, XIX, 125-130; 183-186.
- Callantine, M. F. and Warren, J. M. Learning sets in human concept formation. Psychol. Rep., 1955, 1, 363-367.
- Crafts, L. W. Routine and varying practices as preparation for adjustment to new situations. Arch. Psychol. N. Y., 1927, 14, No. 91.
- Dashiell, J. F. An experimental isolation of higher level habits. J. exp. Psychol., 1924, 7, 391-397.
- Maltzman, I., Bogartz, W., and Breger, L. A procedure for increasing word association originality and its transfer effects. J. exp. Psychol., 1958, 56, 392-398.
- Morrisett, L. and Hovland, C. I. A comparison of three varieties of training in human problem solving. J. exp. Psychol., 1959, 58, 52-55.
- Robinson, Helen, M. Why pupils fail to read. Chicago: The University of Chicago Press, 1946.
- Rothkopf, E. Z. Stimulus similarity and sequence of stimulus presentation in paired-associate learning. J. exp. Psychol., 1958, 56, 114-132.
- Soffietti, J. P. Why children fail to read: a linguistic analysis. Harvard educ. Rev., 1955, XXV, 63-84.

The Learning of Variable Grapheme-to-Phoneme Correspondences:

Variations in the Initial Consonant Position

Harry Levin and John Watson

Cornell University

In a previous study, the writers tested the hypothesis that training on a variable list of words--one in which a single graphic element is associated with two phonemes--would lead to more efficient learning of subsequent variable lists than would training on a list in which each grapheme corresponded to a single sound. In addition, we hypothesized that the variable list would in itself be acquired more easily than the constant one. The first hypothesis was predicated on theory that variable training would create a "set for variability" which yield positive transfer to new learning in which such an "attitude" would be useful. The second hypothesis was based on the notion that response variations would facilitate intralist discriminations on which effective association learning is based.

The findings in the earlier study generally confirmed the first hypothesis but was inconclusive in regard to the second. In that study the grapheme-to-phoneme variations occurred on the medial vowel of tri-phoneme words. An analysis of the intralist confusions indicated clearly that the major source of confusion errors occurred when two or more words in a list shared an initial grapheme and phoneme; next, confusions were due to the sharing of terminal elements; and least errors of this type occurred when words shared only the medial elements.

These findings led to a new experimental variation and to an additional research problem, both of which are pursued in the present study. Since the major cues in association learning appear to be the initial elements in the words, variations in this position should create a stronger set toward variability than did the variations in the medial position which is a relatively unimportant cue. Secondly, the earlier study did not permit us to determine whether the source of confusion was due to the fact that several words in the list shared common visual stimuli or verbal responses or a combination of these factors.

The present study, therefore, follows the earlier one in essential details, although the design is somewhat more elaborate. The major change is that under some conditions the initial grapheme stands for one sound and under other conditions, is associated with two phonemes.

METHOD

Subjects: The Ss were sixty girls attending a Girl Scout Camp near Ithaca. Their average age was 10.2 years, with a range of nine to twelve years.

Procedure: The study used a 2 x 2 factorial design. The sixty Ss were divided randomly into two groups, one of which learned a constant list and the other a variable one. Then, the 30 Ss in each group were randomly placed in two groups, one of which learned a constant and the other a variable list. The ensuing four groups with 15 Ss in each may be labelled as follows: C₁C₂, C₁V₂, V₁C₂, V₁V₂. Two experimenters gathered the data and they appear in each group without bias.

Materials: The lists which the various groups learned are given in Table 1.

Table 1. Lists of Words Learned by the Various Groups

C ₁	V ₁	C ₂	V ₂
come	some	gate	gate
calm	calm	gale	jail
cup	cup	gear	jeer
king	sing	goal	goal

These words are the English equivalents of graphic representations which were rendered in an artificial orthography. Each of the words is made up of three phonemes, and is a monosyllable of the form: consonant-vowel-consonant. Other than the initial grapheme in the variable lists, each grapheme was associated with a single phonemic sound. Two sets of graphic symbols were used: one for the C₁ and V₁ lists and another for the C₂ and V₂ lists.

The lists were matched for frequency of occurrence in English according to the Lorge-Thorndike lists. Lists C₁ and V₁ are matched, as are lists C₂ and V₂, although the second set was somewhat rarer and presumably more difficult than the first.

As in the previous study each word was lettered on a card which was placed on a stand in front of S. E gave the associates on the first trial and then S made the associations to a criterion of three successive correct responses to the total list or until he reached 15 trials.

The instructions to Ss and other details of procedure were the same as in the earlier study (Levin and Watson, 1961).

RESULTS

Since the two scores, trials to criterion and total number of errors, are highly correlated ($r = .92$), only the latter scores were analysed. The error scores were normalized by a log transformation to correct the negative skewness in their distribution.

Comparison of the C_1 and V_1 lists. One of the hypotheses carried over from the first study was that variable lists are more easily acquired than constant lists. This prediction may be tested by comparing the error scores of the 30 Ss who learned the C_1 list with the other half of the sample who learned the V_1 list. These two groups were exposed to the same visual stimuli in the two lists and differed only in the verbal associates given to the two lists. Actually, as can be seen in Table 1, the groups shared two visual-verbal associates and differed only in learning the words, some, sing, calm and king. The mean number of errors made on the C_1 list was 6.50 and on the V_1 list, 10.73. These two means are significantly different, $t = 2.27$, $p < .05$.

This finding suggests that our original hypothesis is false. Given simply the task of learning a constant or a variable list, with no relevant prior experiences, a constant list is easier to acquire. Our contrary prediction was based on the assumption that variable responses within a list would facilitate discrimination among the elements in the list, which in turn would make the list easier. Apparently, the task of making varying associations to a single stimulus is harder than the single phoneme to grapheme association. This finding, however, does not preclude that the variable prior experience which we now know to be more difficult, may yield positive transfer effects to subsequent learning.

The transfer effects of constant and variable training. The formal relationships of the second C and V lists to each other are the same as for lists C_1 and V_1 described above. The error scores (transformed to logs) on the second lists were subjected to an analysis of variance according to the nature of the prior training and the nature of the second lists. The means and standard deviations are given in Table 2.

Table 2. Means and Standard Deviations of Errors on Second Lists (N=15 in each group)

	Groups			
	C_1C_2	C_1V_2	V_1C_2	V_1V_2
M	8.07	8.13	8.60	3.27
S.D.	6.80	5.30	7.57	2.54

In the analysis of variance, only the main effect of training condition is significant ($F = 4.01$, $p = .05$, 1 and 56 df). The means would indicate a significant interaction since the obviously deviant error scores are made by the Ss who were trained on V_1 and then learned V_2 . The F test of the interaction between the nature of the first and second lists is 2.85, which is significant at the 10% level. As can be seen in Table 2, the variation within cells is quite large except for the V_1V_2 group.

Our earlier study involved only two groups: C_1V_2 and V_1V_2 . The second group made fewer errors than the first. This finding may be replicated by comparing the two relevant groups in the present study. The means, 8.13 and 3.27 differ significantly ($t = 3.10$, $p < .01$) thus

making us more confident in the original finding, although the variation in the list is now on the initial whereas formerly it was on the medial position.

An alternative explanation of the finding regarding the effects of variable training exists. Since there were more errors and more trials in learning the V_1 compared to the C_1 lists, the subsequent effects of variable training may be due to the more "warm-up" training. On the face of it this alternative appears unlikely because of the differences in error scores between the V_1C_2 and the V_1V_2 groups. However, since the interaction effect between first and second lists was only marginally significant, we performed an analysis of covariance on the error scores on the second lists, equating the performance on the first lists.

Table 3. Summary of Analysis of Covariance of Error Scores on the Second Lists

Source of Variance	SS of Errors of Estimate	df	MS	F	P
Total	80,317	58			
Within	67,297	55	1,224		
Adjusted means	13,020	3	4,340	3.55	<.05

There is a significant variation among the four groups irrespective of the longer training on the original V list.

In summary, the results indicate that, when Ss are given an initial task of learning a constant compared to a variable list, the latter is more difficult. On the other hand, the original variable training makes subsequent variable lists easier to acquire. This effect is quite specific, since the variable training has little effect on a following

constant list, and original constant-list training has little differential effect on later constant or variable list learning. It may also be noted that the effects of the pre-transfer training override the first finding that a constant list is easier to learn than a variable one.

Confusion Errors

Intralist confusion errors, according to our earlier study, are due almost entirely to the common elements which words in the list share. In that study we isolated four types of confusion errors: Type I, initial elements shared; Type II, terminal elements in common; Type III, the first element of one word is the terminal in another; Type IV, no or only medial elements are shared. Since in the present study, all words within a list share the initial graphic symbol, it was not possible to isolate Type I errors as defined above, nor did the lists make possible Type III errors. Besides, the two V lists vary the one-to-one correspondences between the grapheme and the phoneme. First the confusion errors made to the constant lists will be presented because they provide the closest replication of the first study.

The error types in Table 4 are as follows:

- Type I. Words share initial and terminal graphemes and phonemes.
- Type II. Words share initial and medial graphemes and phonemes.
- Type III. Words share only initial graphemes and phonemes.

The 30 Ss who learned the C_1 list made a total of 157 confusion errors. The mean frequency of errors, as predicted, is highest for Type I and lowest for Type III. The finding is the same for the 194 errors made to the C_2 list. Although the conditions of the present experiment precluded a direct replication of the first study, the elements in common do yield congruent results.

Table 4. Analysis of Confusion Errors on the Constant Lists

Stimulus	Class of Confusion Errors		
	I	II	III
	List C ₁		
come	28 calm	22 cup	2 king
king			6 6 11 come, calm, cup
calm	35 come		12 3 cup, king
cup		18 come	8 6 calm, king
Totals	63	40	54
Mean	31.50	20.00	6.75
Mean %	54.08	34.33	11.59

	List S ₂		
gale	53 goal	18 gate	9 gear
gear			11 20 5 gale, goal, gate
goal	36 gale		15 12 gate, gear
gate		10 gale	1 4 goal, gear
Totals	89	28	77
Mean	44.50	14.00	9.63
Mean %	65.32	20.55	14.13

Confusions of sight or sound? From the information in the earlier study and the error analysis presented above we are not able to say whether the confusion errors are due to the common graphic stimuli or to the sounds common to the two words. The analysis of confusion errors made to the two variable lists, however, permit us to answer this question definitively. In list V_1 , for example, confusions based on common graphic elements alone would predict a low order of errors between the words, some and sing. On the other hand, since these two words share a common phoneme in the initial position, we predict that if the confusions are phonological there will be a higher order of intersubstitutions of these two words.

The confusion errors to the V_1 list are presented first in the order predicted by common graphic elements, and, in the lower part of the table, the order is the one predicted on the basis of common sounds. List V_2 is presented in the same way in Table 6. In the phoneme confusions sections of the tables, the error types are as follows.

- Type I. Words share initial sounds.
- Type II. Words share terminal sounds.
- Type III. Words share medial sounds.
- Type IV. Words share no sounds.

For both lists, the errors follow the order predicted by the common visual stimuli, and are not predictable from the sounds that the words share. Consider, for example, some of the confusions in list V_1 , Table 5. The word some shares a terminal grapheme and phoneme with calm, and an initial phoneme and grapheme with sing. In this list initial graphemes do not differentiate among words. There were 59 confusions between some and calm and only 12 between some and sing.

Table 5. Confusion Errors Ordered According to Grapheme and Phoneme Predictions: List V₁

Stimulus	Class of "Grapheme" Confusions		
	I	II	III
some	59 calm	28 cup	12 sing
sing			15 8 11 some, calm, cup
calm	31 some		13 4 cup, sing
cup		29 some	4 6 calm, sing
Total	90	57	83
Mean	45.00	28.50	10.38
Mean %	53.65	33.98	12.37

	Class of "Phoneme" Confusions			
	I	II	III	IV
some	12 sing	59 calm	28 cup	
sing	15 some			8 11 calm, cup
calm	13 cup	31 some		4 sing
cup	14 calm		29 some	6 sing
Total	54	90	57	28
Mean	13.50	45.00	28.50	7.25
Mean %	14.32	47.75	30.24	7.69

Table 6. Confusion Errors Ordered According to Grapheme and Phoneme Predictions: List V₂

Stimulus	Class of "Grapheme" Confusions		
	I	II	III
jail	37 goal	15 gate	3 jeer
jeer			6 11 1 jail, goal, gate
goal	13 jail		11 8 gate, jeer
gate		11 jail	2 1 goal, jeer
Total	50	26	43
Mean	25.00	13.00	5.37
Mean %	57.64	29.97	12.38

	Class of "Phoneme" Confusions			
	I	II	III	IV
jail	3 jeer	37 goal	15 gate	
jeer	6 jail			11 1 goal, gate
goal	11 gate	13 jail		8 jeer
gate	2 goal		11 jail	1 jeer
Total	22	50	26	21
Mean	5.50	25.00	13.00	5.25
Mean %	11.28	51.28	26.67	10.77

In the same list, calm and some is the predicted highest graphic confusion and calm and cup is the phonic confusion. In this comparison, the graphic confusion is more potent, and this is true for all of the individual comparisons in list V_1 and V_2 . We may conclude, therefore, that the basis for the confusion errors is the visual stimulus rather than the verbal response.

SUMMARY

This study tested the propositions that a list containing variable grapheme-to-phoneme correspondences is easier to learn than a list in which the associations are constant and that training on the variable list would facilitate subsequent variable list learning. The evidence for the first hypothesis is clearly contrary to the prediction: given no prior experiences with such lists of words, constant lists are acquired with fewer errors.

The second prediction was confirmed: most efficient learning occurred when two variable lists followed each other, compared to the conditions of variable or constant lists following an initial constant list, or a constant list after prior learning of a variable one.

An analysis of intralist confusion errors indicates that errors are due to words sharing common graphic stimuli rather than to common sounds.

REFERENCES

- Levin, H. and J. Watson. The learning of variable grapheme-to-phoneme correspondences. Mimeo, 1961, 23 pp.

THE LEARNING OF VARIABLE GRAPHEME-TO-PHONEME CORRESPONDENCES:

COMPARISON OF ENGLISH AND SPANISH SPEAKERS

Harry Levin, Esther Baum and Susan Bostwick

Cornell University

Several previous studies (Levin & Watson, 1961a,b) have indicated that when a graphic element is associated with two verbal associates, positive transfer to subsequent variable associations is facilitated. This result is true also for problem solving (Ackerman & Levin, 1958, Morrisett & Hovland, 1959). Recently, Ianco (1963) has demonstrated that a concurrent bilingual model of vocabulary learning, in which two verbal associates to the same graphic stimulus are learned in the same list, yields superior positive transfer to a condition in which the two lists are learned consecutively. We have hypothesized that the positive transfer is a consequence of the acquisition of a "set for diversity" which leads the child to expect new associations to old stimuli. Another explanation of these findings runs like this: variable associations create greater differentiation among items within the list which, in turn, lead to positive transfer (Ianco, 1963).

A study with Italian children, however, gave results completely contrary to the earlier findings. For them, constant lists were most beneficial to transfer. We have speculated that these unexpected results are a function of the Italian language. The grapheme-to-phoneme correspondences in Italian are quite regular compared to the variations in English. All of our Italian subjects were 10 to 12 years old and literate. Learning variable verbal associates to a single grapheme was probably novel to them so that their performance on the experimental task suffered from negative transfer from their strongly habituated single

sound to letter experiences. Perhaps, the set for diversity effects would not have been masked if we had used a nonverbal task.

This study was designed to test directly our hypothesis for the divergence in results between English and Italian speaking subjects. Most available to us were college students whose primary language was Spanish. The variation of grapheme to phoneme correspondences was equally served by Spanish as by Italian. However, the variations in habituated language experiences between the Spanish and English speaking subjects was not as clear as we would have liked. Our Spanish sample were all students at an English speaking university and had had varying degrees of training in speaking and reading English. One might argue that their second language learning has made them especially aware of the variable correspondences characteristic of English. Since all of earlier work had been done with elementary school children, a second purpose of this study was to test the generality of our findings to adults.

METHOD

Subjects

The Ss of this study were 40 native English speaking Americans and 40 Spanish speaking students enrolled in Cornell University during the fall semester, 1962. The Spanish speaking group, composed of students who were either from Latin America or from Spain, consisted of 36 males and 4 females who ranged in age from 18 to 36 years, with a mean age of 26. Ten of the Spanish speakers had studied English since the first grade, 24 had studied English through high school, and five had had no formal training in English prior to their arrival in the United States. The Spanish Ss had been in the U. S. from 2 months to 7 1/2 years. The American sample consisted of 24 males and 16 females, from 18 to 35 years of age with a mean of 22 years.

Materials

The design of the study required that four lists of graphic symbols and their verbal associates be developed (see Table 1). Each list comprises eight nonsense syllables taken from the Glaze list in Underwood and Schulz (1960). The stimulus materials are an artificial orthography used in previous studies (Levin and Watson, 1961).

There are certain relationships between the graphic materials and the nonsense syllables. In the constant lists (labelled C_1 and C_2), each grapheme is related to only one phoneme. In the variable lists (V_1 and V_2) one grapheme in the initial position is used to represent two different sounds; that is, in each variable list, the first grapheme was pronounced one way in half the words and had another value in the other half: /S/ or /K/; /V/ or /N/.

Each S learned two lists: C_1 or V_1 followed by C_2 or V_2 . The first and second lists contain certain similarities to each other. Three of the medial vowels in the first lists appear also in the second; one medial vowel is unique to each set of lists. There is no overlap of terminal consonants between first and second lists.

The graphic stimuli were each mimeographed on 5" x 8" cards which were placed on a small stand before the S.

Procedure

The design is a 2 x 2 x 2 factorial one. Each of the English and Spanish speaking Ss was assigned to one of four conditions defined by the sequence of lists he learned: C_1 followed by C_2 , C_1V_2 , V_1C_2 , V_1V_2 .

The details of instructions and procedures were as follows:

We're interested in finding out how people learn a new language. In order to do this, we've invented both artificial words and artificial letters. Like any language, the sounds of the words and the letters are closely associated and I'd like to see if you can learn the association.

Table 1. Four Lists of Nonsense Syllables and Their Graphic Representations

C ₁	Symbol	C ₂	
kug	山 X r	vud	C X N
kef	山 7 p	vem	C 7 A
kaz	山 F T	vam	C E A
kag	山 E t	vor	C W 8
kif	山 9 p	vop	C W A
kub	山 X 4	vop	C 7 A
kiz	山 9 T	vur	C X 8
keb	山 7 4	vad	C E N
V ₁		V ₂	
sug	山 X r	vud	C X N
kef	山 7 p	nem	C 7 A
saz	山 E T	vam	C E A
kag	山 E t	vor	C W 8
sif	山 9 p	nop	C W A
kub	山 X 4	vop	C 7 A
kiz	山 9 T	nur	C X 8
seb	山 7 4	nad	C E N

Now I'm going to show you each word written in these artificial letters. As I show them to you, I'm going to tell you what they represent. I'd like you to try and learn these representations. This first time just repeat each word after me as I show it to you, and try to pronounce it exactly as I say it.

Each card in the first trial was exposed for a 10 second interval.

Now I'm going to show you the same written forms in a different order. This time I want you to try and tell me what each one represents, then I'll say it after you. It's perfectly all right if you guess.

After the subject had responded, regardless of whether he had guessed correctly, or after the 10 seconds were up, the correct association was given by E. Each card was exposed for the full 10 seconds, even if the subject guessed immediately. The order of stimuli were randomized for each trial so that serial effects would not appear.

The criterion for learning was 3three successive errorless trials, but if criterion was not reached by the 15th trial, testing was terminated for that particular list. Immediately after Part I was over, Ss were told:

Now I've got another set of words for you to learn.
We'll do it in the same way as we did the first set.

Essentially the same procedure was followed in Part II as in Part I. Total testing time was anywhere from 20 to 75 minutes, depending on the combination of lists.

All responses made by the subject to any given stimulus were noted, so that an analysis of the errors could be made later.

RESULTS

The two scores gathered in this study were the number of trials to reach the criterion of learning and the number of errors. We have chosen to base our analyses on number of errors, which makes this study

comparable to the earlier ones.

In Table 2 are presented the mean number of errors for the various lists. Our first hypothesis states that lists containing variable associations would be more difficult for the Spanish group, because of their lack of experience with this condition, than for the English speaking Ss. The Spanish Ss made more errors on both the V_1 and V_2 lists than did their American cohorts. The difference approaches statistical significance for the V_1 list ($t=1.61$, $p<.10$) and is significant for the V_2 list ($t=2.41$, $p<.05$). We cannot attribute these differences to any general learning superiority of one group over the other because the two language groups performed essentially at the same level on the two constant lists.

The transfer effects of learning one or the other types of lists on subsequent learning are measured by the differences in error scores between the two lists. For a variety of reasons, we expect the second lists to be generally learned more easily than the first. Warm-up effects, for example, should enhance the acquisition of the second lists. Also, the second lists should be learned more easily than the first since three of the medial vowels in the first lists appear also in the second. There should, however, be no differences in the warm-up contribution to the various learning scores.

For each S, the savings in learning was indexed by the differences in numbers of errors between the first and second lists. The means and mean differences are given in Table 3. These differences were subjected to an analysis of variance according to the classifications, language of the S, nature of the first list and nature of the second list. The two main effects, type of first list ($F=46.41$, $p<.01$) and type of second list ($F=35.88$, $p<.01$) were significant. Also, the triple

interaction indicates significant variations among the cell means ($F=5.98$, $p<05$). There were no differences in transfer directly attributable to the language groups of the Ss.

Table 2. Mean Errors Made to Each List (Each Mean Based on $N=20$).

	American	Spanish
C_1	36.1	30.5
V_1	52.7	65.0
C_2	6.9	6.6
V_2	26.9	43.6

Table 3. Mean Errors and Savings in Errors (Each Mean Based on $N=10$)

Condition	American		Spanish	
	Means	Mean Difference Trial 1 - Trial 2	Means	Mean Difference Trial 1 - Trial 2
C_1	38.4	29.5	21.4	19.5
C_2	8.9		1.9	
C_1	15.0	-9.0	39.5	-2.8
V_2	24.0		42.3	
V_1	43.5	38.6	66.3	55.0
C_2	4.9		11.3	
V_1	62.1	32.3	63.7	18.8
V_2	29.8		44.9	

These sources of significant variations in the data may be described in the following ways. The savings, or positive transfer, is greater for those Ss who initially learned a variable compared to a constant list. Also, the savings are greatest when the second list is a constant one. The latter finding is contributed mainly by condition V_1C_2 , because the C_2 list seems particularly easy. More will be said about this in the following discussion.

Both groups showed positive transfer from the V_1 to the V_2 list. However, the mean savings was 32.3 errors for the English speaking group and 18.8

for the Spanish Ss. This difference approaches significance ($t=1.78$, $p<.10$). The one condition showing a loss is the constant condition followed by the variable one.

DISCUSSION

This study was designed to resolve different sets of results from several studies. Two experiments with English speaking children indicated that training in variable grapheme-to-phoneme correspondences, although more difficult to acquire than constant correspondences, resulted in more positive transfer to subsequent association learning. The same study with Italian children indicated that constant association yielded higher transfer. This discrepancy was attributed to the characteristics of the languages and writing systems of two sets of Ss.

To test the hypothesis that variable learning is more difficult for speakers whose native languages have consistent correspondences, the study was repeated with Spanish speaking and English speaking adults. Variable lists are clearly more difficult for the Spanish speakers, although constant lists were equal in difficulty for the two language groups. Therefore, this much of the hypothesis is confirmed.

The transfer effects are complicated to understand. In the study with American children (Levin and Watson, 1961b) the only condition which yielded substantial improvement was one variable list to another variable list. The Italian study, on the other hand, indicated that initial training on the constant list was most conducive to both subsequent variable and constant learning, especially to the latter. Our present study may best be summarized as "positive transfer is a function of the similarity of the two tasks."*

* At first glance, the large savings from the V_1 to the C_2 condition is impressive. We believe, however, that this result is a function of the difficulty of the V_1 list and the ease of the C_2 list, independent of each other. Of course, to pin down this point we should have a control condition in which the C_2 list is learned without a prior list.

The present study differs from the others in two possibly important respects: the Ss are adults rather than children, and the responses were nonsense syllables instead of common words. We expect that the variable-constant distinction may be more easily inferred from familiar words than from nonsense materials.

The series of studies, of which this is one, started with some speculations about the process of learning to read English. Specifically, we have been concerned with the learning of grapheme-to-phoneme correspondences which presents certain problems for English. Since a given grapheme may be decoded as more than one sound, one pedagogical suggestion has been to teach children an artificially regular spelling to sound system and then to introduce irregular forms. We have questioned this procedure on the grounds that, though initial acquisition of correspondences may be facilitated, the transfer to the decoding of new instances may be retarded. The multiple correspondences of English is an inevitable condition which the reader must face and learn to handle. Given this, the results of the present study indicate that initial training in variable correspondences yields positive transfer to new variable forms. While it is true that initial constant training yields positive transfer effects to new constant forms, this is simply not the conditions for reading English. Our question must be, What conditions facilitate the learning of variable correspondences? The results from this and earlier studies indicate that initial variable training, compared to a constant one, is considerably superior.

SUMMARY

This study follows on several others which found that children trained to learn two verbal associates to a single graphic stimulus generalized

this skill to the learning of subsequent variable associations. This learning problem is designed as an analog to learning grapheme-to-phoneme correspondences as a part of the process of learning to read. However, the study repeated with Italian children indicated that the variable condition was difficult to learn and only the constant condition yielded positive transfer effects. In this study, we tested the hypothesis that variable associations were difficult for readers of a language which had a low instance of variable associations. Forty Ss whose native language was Spanish and forty speakers of English learned one of four combinations of lists: constant-variable, C-C, V-C, V-V. Variable lists were somewhat more difficult for the Spanish than for the English speaking as Ss. Both initial training on constant and variable lists yielded a savings in learning the second lists of the same type. We have argued that since learning to read English requires variable associations of the type that we have studied, it would be well to train initially on variable instances.

REFERENCES

- Ackerman, W. I., and Levin, H. Effects of training in alternative solutions on subsequent problem solving. J. educ. Psychol., 1958, 49, 239-244.
- Ianco, Anita D. Some transfer effects of concurrent and consecutive bilingualism. Unpublished Master's thesis, Cornell University, 1963.
- Levin, H., and Watson, J. The learning of variable grapheme-to-phoneme correspondences. Mimeo, 1961a, 23pp.
- Levin, H., and Watson, J. The learning of variable grapheme-to-phoneme correspondences: variations in the initial consonant position. Mimeo, 1961b, 12 pp.
- Morrisett, L., and Hovland, C. I. A comparison of three varieties of training in human problem solving. J. exp. Psychol., 1959, 58, 52-55.
- Underwood, B. J., and Schulz, R. W. Meaningfulness and verbal learning. Philadelphia: Lippincott. 1960.

THE USE OF CUES IN WORD RECOGNITION

Gabrielle Edelman ¹
Cornell University

The basic problem with which this study is concerned is the visual recognition of words by non-readers and beginning readers. When a child is confronted with a word he has never seen and is asked to remember it he must utilize some identifying mark, i.e., a cue. The purpose of this experiment is to discover which cue or cues are the most important. The following questions will be asked:

1. What are the cues in a word by which non-readers and beginning readers remember that word?
2. Are the same cues utilized in recognizing a long word and a short word?
3. Do non-readers and beginning readers use the same cues?
4. Do boys and girls use the same cues?

METHOD

A delayed-recognition task was constructed with trigrams and five-letter paralogues. S was given the opportunity to recognize these nonsense words on the basis of any one letter cue, or on the basis of shape (shape was defined as whether the letters were above, on, or below the line in lower case).

Ss were tested individually; each S was first shown, for 5 seconds, a 4 x 6 inch white card on which a nonsense word was typed (1/4 inch high for letters "on" the line) in lower case. Then the card was

1. This is an abbreviated version of a thesis submitted to Cornell University for the M.A. degree. The complete version is available from the Cornell University Library.

withdrawn from sight, and S was shown a second card (the response card) also a white 4 x 6 inch card, on which a group of words were typed with the same size type as the word on the first card. The words on the response card were in random arrangement. S was asked to point to the word he had seen on the first card, or the one most like it.

The words on the response card represented various errors in recognition S might make as a result of the shape of the word or the various letters of the word. For example, if the trigram "cug" was shown on the first card, the choices on the response card were: "arp" (same shape as cug but all the letters are changed), "che" (same first letter as "cug" but the second and third letters and the shape is changed), "tuk" (same second letter as "cug" but the first and third letter and the shape is different), and "ilg" (same third letter but the first and second letters and the shape is changed).

The combinations of four cues in the trigrams (shape, first letter, second letter and third letter) were varied systematically on the response cards, so that each item represented a different set of errors. Every cue was presented in every combination with every other cue. So that all the possibilities could be realized, cues had to be held constant in turn, i.e., kept the same as in the original word on the first card.

The same system was constructed for the five letter paralogues, in which the combinations of six cues (shape, first letter, second letter, third letter, fourth letter, fifth letter) were systematically varied on the response cards.

There were eight combinations (items) in the trigrams, and 52 combinations in the paralogues. Each item was doubled (i.e., an analogous

cue choice with different letters was given in the task) so that reliability could be estimated. All together then, there were 120 items in the task, and they were presented in random order.

SUBJECTS

Ss were 50 kindergarteners (25 boys and 25 girls) and 50 first graders (25 boys and 25 girls) drawn at random from the Dryden Central School in Dryden, New York.

There is no instruction in reading or in the alphabet in the kindergarten. First graders were instructed by whole word and sentence methods. New words were generally introduced in sentence context with the first letter of the new word given as a clue. At the time of testing, the first graders had 5 months of this type of training. They were beginning to read from the McKee preprimers.

RESULTS

Reliability:

Reliability actually represents the consistency of recognition on the basis of analogous cue variations. Average reliability (in percent) is presented in Table 1.

Table 1. Average Reliability in Percent

Kindergarten.....	.54
First Grade.....	.76
Girls.....	.66
Boys	.65

First graders recognized words from the same cues more consistently than kindergarteners.

Task results:

A frequency count was made of the number of times Ss recognized words on the basis of certain cues. The totals are presented in Table 2.

Table 2. Totals of the Number of Times Ss Recognized on the Basis of Each Cue

Cues:	Five Letter Words						Trigrams			
	Shape	First	Second	Third	Fourth	Fifth	Shape	First	Second	Third
Kinder-										
garten	330	1673	540	755	518	1365	69	326	140	265
1st Gr.	80	2497	757	693	287	930	20	451	146	176
Boys	214	1886	589	780	445	1208	43	358	143	253
Girls	196	2284	708	614	360	1087	46	419	93	188

The first letter of both the long and short word forms was the cue most utilized by non-readers and beginning readers in word recognition. The last letter is also an important cue, especially for kindergarteners, and more so in trigrams than in long words.

Statistical analysis using Friedman's two-way analysis of variance in a comparison of all six cues in the paralogues reached .001 significance level for each group (i.e., the four groups, kindergarten boys, kindergarten girls, first grade boys and first grade girls were calculated separately), demonstrating the fact that the differences in number of times Ss recognized paralogues on the basis of different cues was significant. A comparison of the four cues in the trigrams also reached .001 significance level for each group.

The Cochran Q Test was used to compare one cue with another cue. In a comparison of the two strongest cues, i.e., the first and last letters, the first letter was shown to be utilized significantly more times than the last letter in all groups except the kindergarten boys, for whom a great deal of competition between the first and last letters exists in word recognition.

Shape was shown to be significantly weaker than even the weakest letter cue for all groups in both long and short words.

A summary of all the statistical comparisons is shown in Table 3.

TABLE 3

Significance levels reached by each group using Friedman's two-way analysis of variance

<u>Comparison of:</u>	<u>Kindergarten</u>		<u>First Grade</u>	
	<u>Boys</u>	<u>Girls</u>	<u>Boys</u>	<u>Girls</u>
S,1,2,3,4,5 (5-letter)	.001	.001	.001	.001
S,1,2,3 (3-letter)	.001	.001	.001	.001
S,2,3,4 (5-letter)	.001	.001	.001	.001
2,3,4	.001	.001	.001	.001

Significance levels reached by each group for each comparison, using the Cochran Q-Test

<u>Five Letter Comparisons</u>	<u>Kindergarten</u>		<u>First Grade</u>	
	<u>Boys</u>	<u>Girls</u>	<u>Boys</u>	<u>Girls</u>
1 & 2	.001	.001	.001	.001
1 & 3	.01	.001	.001	.001
1 & 4	.001	.001	.001	.001
1 & 5	.7	.01	.01	.001
1 & S	.001	.001	.001	.001
2 & 3	.001	.7	.7	.05
2 & 4	.8	.7	.01	.001
2 & 5	.001	.001	.1	.7
2 & S	.01	.01	.001	.001
3 & 4	.001	.01	.001	.001
3 & 5	.001	.001	.2	.8
3 & S	.001	.001	.001	.001
4 & 5	.001	.001	.001	.001
4 & S	.01	.1	.001	.01
5 & S	.001	.001	.001	.001

Three Letter Comparisons

1 & 2	.01	.001	.001	.001
1 & 3	.2	.2	.1	.001
2 & 3	.05	.01	.2	.2
2 & S	.001	.05	.001	.001
3 & S	.001	.001	.001	.001

INTERPRETATION OF RESULTS

If no one cue was more important than any other cue in recognizing words, Ss would have recognized the words by "chance" cues, i.e., no single cue would have been used more than any other cue, and Ss would tend to recognize the words by one cue one time, and by another cue another time. As can be seen from the frequency data and the subsequent analysis, word recognition on the basis of chance cues is most unlikely.

In the five letter series the strongest cue (the first letter) was significantly stronger than the second strongest cue (the fifth letter for all groups except first-grade girls, whose second strongest cue was the second letter), except for kindergarten boys. Thus, the first letter is the most important cue for non-readers and beginning readers in word recognition of a long word (i.e., five letters). Because many of the kindergarten boys recognized five letter paralogues on the basis of the fifth letter as well as the first letter, the comparison of the cue power of these two cues was not significant for this group. Thus, among male non-readers a great deal of competition exists between the first and fifth letters, the two most important cues in word recognition of paralogues.

In the trigram series the first letter is the most important letter for all groups; and the third letter is second most important for all groups except the first-grade girls, for whom the second letter is next in importance. Nevertheless, a statistical comparison of the two strongest cues in the three letter series did not show significance for any group except the first-grade girls. This indicates that competition between the two most important cues in the trigram series was strong, and both the first and third letters are very important for word

recognition among kindergarteners and first-grade boys. The high level of significance reached by the first-grade girls in this comparison shows that for them, the first letter is certainly more important than any other cue in a trigram.

The cue value of shape is the weakest for all subjects, and shape is a significantly weaker cue in word recognition than the weakest letter (except for kindergarten girls in the longer series).

QUALITATIVE PATTERNS

Although enough Ss followed the pattern of utilizing the first letter as the most important cue and the last letter as second in importance to yield statistical significance, there were some Ss who did not follow this pattern. Ss with special qualitative patterns were mainly among the kindergarten boys, and the first-grade girls.

The kindergarten boys often utilized the last letter of a word as their most important cue, and the first letter as the second most important cue. This may be because non-reading boys tend to scan in the opposite direction, i.e., from right to left in a word, rather than from left to right. Or perhaps they scan from left to right but tend to remember the most recent cue (i.e., the last letter) better than the first cue they noticed.

The first grade girls had among them many who utilized the first letter as most important, but the second letter as next important, the third letter as third important, etc., in both the trigrams and paralogues. These Ss were noted at the time of testing because they did not seem to depend wholly on visual cues in word recognition. These subjects tended to say, out loud, the names of the letters in the original word in order to help them remember the word. And they would say the letters in order,

from left to right. Then, on presentation of the response card, they would search for the letters in that order; often by the time they had reached the end of the word, they had forgotten the last letters, especially in the five letter series.

There seems to be little doubt that these latter Ss are spelling. The theory of recency certainly does not apply to Ss who spell; rather a theory of letter order applies. These Ss attempted to remember all the letters of a word by saying them aloud from left to right, and then in the process of recognition, scanning the response card for the letters in the order in which they tried to remember them, i.e., from left to right.

THEORETICAL IMPLICATIONS

Theories which propose that non-readers and/or beginning readers recognize words as wholes by their shape have not been supported in this study. The shapes of words, offered as a cue next to letter cues, was rejected in favor of letter cues by Ss in this experiment.

This study has supported theories which propose individual letters as cues used by non-readers and beginning readers. Furthermore, this study has supported the proposal that the first letter in particular, and also the last letter of a word are the most important cues used by Ss who are not very familiar with the alphabet.

An explanation of this "first-last" letter phenomenon may lie in the theory of primacy and/or recency of cues.

Or perhaps the first and last letters of a word stand out in particular in the child's imagery because they are isolated on one side by a white space whereas middle letters are embedded with letters on either side.

The first grade Ss who recited the letters of the original word in order to remember them were not utilizing the same cues as Ss who did not depend on oral recitation. These first graders obviously show the effects of instruction--perhaps home instruction in the alphabet. These Ss knew the names of the letters they saw well enough to use a verbal mediator as a cue, instead of, or in addition to, a strictly visual one.

PRACTICAL IMPLICATIONS

The major implication for the teaching of reading is that the basic belief on which the whole word method of teaching reading lies (i.e., the belief that children recognize words by their shape) is incorrect. Educators may believe the child is attending to the whole word, when he is actually utilizing certain letter cues. Helping pupils learn the letters well so that they may use letter cues to the best of their ability would be an important teaching improvement.

Furthermore, since the first and last letters are the cues most often used by untrained children, teaching concentration should be placed on the middle letters of words, so that children may learn to distinguish words on the basis of other letter cues besides the first and last letters, which they already do without instruction.

CONCLUSIONS

1. Letter cues, and not shape cues are the bases by which kindergarteners and first graders recognize words.
2. The first letter of a word is the most important cue in word recognition.
3. The last letter of a word is used as the second most important cue in word recognition except for first graders who know the alphabet very well; they tend to use the letter cues in their order from left to

right, the first being the most important cue, the second letter next in importance, etc.

4. Some kindergarten boys utilize the last letter as the most important cue in a word, and the first letter as the second most important cue.
5. In short three-letter words, the last letter is a more important cue than it is in longer, five-letter words.
6. The implications of these conclusions for teaching reading are that the letters ought to be taught more thoroughly so that beginning readers might utilize all the letter cues in a word to their advantage, rather than just the first and last letters, which they apparently do without instruction.

¹This research was supported by a grant from the U.S. Office of Education, Department of Health, Education and Welfare. Appreciation is expressed to Dr. Florence B. Minicuti and Etta Woodward for data collection, to Etta Woodward for assistance in construction of stimulus materials, and to Mabel Lam, Marion Prentice and Linda Jordan for assistance in statistical analyses.

GEOMETRIC FORM AND DETAIL AS DETERMINANTS OF COMPARATIVE
SIMILARITY JUDGEMENTS IN YOUNG CHILDREN¹

Henry N. Ricciuti
Cornell University

The present study grew out of the writer's interest in the multiplicity of stimulus cues to which young children respond when asked to group objects, or to make comparative similarity judgements, as on matching-with-standard tasks. A familiar generalization from early studies of concept formation or sorting behavior states that normal adults and older children categorize stimuli, or judge their similarity, primarily on the basis of "abstract" or representationally meaningful class concepts, while disturbed or retarded adults, and young children tend to be more responsive to "concrete", or purely "perceptual" properties of stimuli, which tend to be viewed as non-relevant cues by the average adult (e.g. 8, 15). In recent years, it has become increasingly apparent that this highly gross distinction between "abstract" and "concrete" modes of thinking or perceiving, with its implicit "good-bad" connotation, has been considerably less than useful in furthering our knowledge about the processes of perceiving and categorizing stimuli, particularly from a genetic point of view. A number of recent investigations have placed primary emphasis upon more precise

¹This research was supported by a grant from the U.S. Office of Education, Department of Health, Education and Welfare. Appreciation is expressed to Dr. Florence B. Ricciuti and Etta Woodward for data collection, to Etta Woodward for assistance in construction of stimulus materials, and to Mabel Lum, Marion Prentice and Linda Johnson for assistance in statistical analyses.

identification of the variety of different ways in which adults and children of various developmental levels categorize stimuli, with these different modes of categorization being regarded as representing individual "styles of conceptualization", whose antecedents and consequences in the development of the individual need to be systematically ferretted out (e.g. 5, 9, 10). At the same time, it has become increasingly clear that in addition to our growing knowledge of individual, group, or developmental differences in "styles of conceptualization" as characteristics of subjects, we need much more systematic information about the perceptual-cognitive processes involved in categorizing behavior, particularly with regard to the role played by various stimulus cues or properties (e.g. 3, 7, 14).

The investigation reported here was one of several studies concerned with developmental changes in the kinds of stimulus properties utilized by children when categorizing or making comparative similarity judgements on groups of stimuli. The purpose of this particular study was to investigate the relative influence of over-all shape and of small details as stimulus cues, in a task situation where either type of cue may serve as a basis for judging which of four simple geometric figures was most like a standard figure. Attention was directed both to developmental changes as well as to the scope and correlates of individual differences in cue utilization.

GENERAL THEORETICAL BACKGROUND

Two closely related processes play a particularly important role in the early perceptual and cognitive development of the young child. On the one hand, the child gradually develops the capacity to distinguish between different objects or stimuli which are previously not discriminable;

in a sense, he becomes increasingly aware of and responsive to differentiating cues or properties of objects. On the other hand, from a very early age, children begin to manifest increasing responsiveness to, or awareness of, similarities between objects or stimuli in the natural world. With continued development and experience, the child becomes aware of the fact that many of the objects in his natural environment are both alike and different in various ways, and that it is sometimes the similarities that are in some sense more important, sometimes the differences.

In the first few years of life, many of the particular differences and similarities the child responds to are more or less discovered by him in the course of play and exploration of the world--these are probably largely determined by the comparative strength or intrinsic "attractiveness" of particular stimulus properties (such as complexity, contrast, size, etc.). Very early, however, as a result of experience and learning, and particularly as a function of language acquisition,² children learn that certain differences and certain similarities are somehow more important or "salient" than others, and that similarities between particular objects are much more important than obvious differences between them. Thus children gradually learn to react selectively to the kinds of similarities and differences which facilitate their making the kinds of identifications and discriminations most demanded of them, or most "salient" for them.

In a given experimental or real life situation, the particular

² The question of the extent to which the language structures or influences modes of thinking is the familiar problem posed by the Whorfian hypothesis (16), which will not be discussed here.

stimulus cues which the child utilizes as a basis for judging that two or more objects are alike or belong in the same category, will depend in large part upon the nature of the task posed for him. Where a specific identification or categorization is required of him (e.g. "which are the large ones", "which are the cars", "which are the tools", etc.), if he understands the criterial attributes of the class of objects designated by the term or terms used, he should select objects on the basis of the relevant stimulus properties, provided he is also able to ignore differences not relevant to the particular identification demanded. On the other hand, if the child is asked simply to categorize a group of objects into sub-sets on whatever basis seems most appropriate to him, or if he is asked which of several possible stimuli is most like a standard, then he has to decide for himself which of several possible identifications or categorizations seems most salient or appropriate. It is precisely for this reason that many investigators, including the present one, have been particularly interested in studying this type of categorizing behavior, since it reveals which of various possible stimulus cues or cue combinations appear most salient to the child as criteria for classifying a set of objects, and what kinds of differences will be ignored in producing such classifications.

In a given test situation of the sort just described, there are at least three major factors which play a role in determining the particular stimulus properties or combination of properties that will be selectively utilized by subjects as a basis for categorizing stimuli.

- a) The intrinsic "attractiveness" or strength of the competing stimulus cues involved in the task certainly represents an important factor to be considered, (even though it is extremely

difficult methodologically to obtain satisfactory measures of competitive cue strength). Particularly in the early years of life, there seem to be a wide variety of stimulus properties to which children attend, and on the basis of which two objects may be seen as similar. We need to know much more about such cues, and how responsiveness to them changes with increasing experience and development.

b) The particular cues responded to selectively by a child will also depend in large part upon the kinds of identifications and categorizations which have been stressed in the child's culture or sub-culture. As already suggested, in the normal course of events, certain cues or object properties become generally more salient or prepotent with increasing age for most children, through social learning and language acquisition, or because they provide the basis for identifications or recognitions which are in some sense functionally or adaptively important in a particular culture. (See also Bruner's discussion of factors influencing "category accessibility", 3, and Gibson's discussion of "distinctive features in perception", 6).

c) Response choices must also be regarded as reflecting characteristics of the individual subject. These characteristics include such factors as developmental level and sex, as well as characteristics which may account for a subject responding to stimulus cues or combinations different from those found to be "modal" for his particular age or cultural group.

There is much still to be learned about the role of all three types of response determinants just mentioned, as they influence comparative

similarity judgements, or categorizing behavior. Certainly we need more information about the dimensions or criteria of similarity which stand out in the very young child's attempt to cognitively organize his world. Further identification and clarification of "modal" identifications or systems of categorizations emphasized by particular cultures or subcultures, and further understanding of the manner in which these become superimposed on the child's "natural" repertoire of identifications and categorizations would certainly be extremely important.³ Finally, determining the antecedents and psychological significance of individual differences in children's categorizing behavior should increase greatly our understanding of the perceptual-cognitive mechanisms involved in such behavior. These are some of the more general concerns which underlie the present study and several which follow. We turn next to a consideration of the more specific background of this study.

BACKGROUND OF PRESENT STUDY AND LITERATURE ON FORM AND DETAIL AS COMPETING STIMULUS CUES

It would be appropriate at this point to indicate why gross form and detail cues were selected as the competing cues to be studied, and to summarize briefly the small amount of directly relevant previous literature. In a previous study of conceptual thinking in young preschool children (11), the writer was impressed with the variety of relatively small, apparently "inconsequential" details which children

³See, for example, Carroll and Casagrande's attempt to demonstrate a relationship between language structure and preferential cue utilization in comparative similarity judgements (4).

pointed out as perceived similarities between pictures of familiar objects, or between simple geometric cutout figures. These details sometimes provided the only apparent basis for matching a choice figure with a standard (as in matching a banana with a spotted dog on the basis of similarity in the small spots on both). In other instances, the details were the only similarities pointed out between two stimuli which most older children or adults would judge as alike because of common membership in some familiar, broad category. In many ways, these detail cues, which many children and adults would regard as irrelevant or non-criterial, resembled the responses described as "concretistic", or "perceptual" by some investigators (e.g. 8, 15). Other evidence of the attractiveness of small details as stimulus properties may be found in the behavior of infants, starting at about ten to twelve months, when they often become intrigued with a tiny spot or other small object clearly outlined against a uniform background, perhaps spending considerable time carefully placing a finger tip on the spot or tiny hole.

In the present study a variety of small details, drawn as parts of simple geometric figures, were "pitted against" over-all stimulus shape as the alternative similarity cue. This was done so as to evaluate the cue value of small details when in competition with that of familiar, simple geometric forms.

Although there have been many studies in which form has been pitted against such competing stimulus cues as color or size, there do not appear to have been any in which the salience of geometric form as a similarity cue has been evaluated against the competing stimulus value of small details. There are several lines of previous research,

however, which have considerable relevance to the type of judgement involved in this research. The many studies dealing in one way or another with the question of "whole versus part" perception, summarized very well by Wohlwill (18), have led to the general conclusion that in the perception of complex forms, young children tend to respond to over-all, global features of the stimulus and have difficulty differentiating small parts. With increasing age, they become capable of identifying parts of the whole, and gradually they are able to deal in a more integrative fashion with both wholistic and detail features of the stimulus. There have been some studies, however, which indicate that young children can readily respond to details when the competing stimulus value of the whole is not too strong in terms of either meaningfulness or complexity.

Whether a particular detail in a figure will be attended to by children will obviously depend in part, at least, on whether it is perceived as an important feature of a meaningful or familiar stimulus. Vurpillot (14) has recently re-emphasized the importance of the fact that very young children often depend upon small details in attempting to identify familiar objects, sometimes erroneously. On the basis of this study, and a previous one with H. Brault (13), Vurpillot concluded that certain special details, which are "characteristic" of the object, play a crucial role in object recognition and in the development of empirical "schema" of familiar objects, particularly with younger children (e.g. the ears of a rabbit, the wool of a lamb, the face of a doll). In the early years, stimuli will not be identified as representing familiar objects unless the required

"characteristic detail" is present. With increasing age, however, the characteristic detail gradually loses its predominance and the child may identify the object on the basis of other features which are present, even though the characteristic detail is missing.

The perception of similarities in drawings of familiar objects primarily on the basis of details in the pictures has recently been considered as indicative of an "analytic" style of conceptualization by Kagan, Moss, and Sigel (9). An experimental measure of analytic style was based on a series of pictorial triads, for each of which the child was asked to select two of the figures which were alike or went together in some way. Responses were scored as "analytic" if similarity was judged on the basis of common details, "relational" if on the basis of some functional relationship between the two choice figures, or "inferential-categorical" if on the basis of common membership in a broad class. The test was constructed to minimize the possibility of inferential responses, since these could be expected to occur with great frequency, so that most responses had to be either analytic or relational. The authors found that under these circumstances, analytic responses tended to increase, relational responses to decrease, from the first to the sixth grade. Individual differences in analytic attitude were reported as showing some degree of consistency across tasks, and promising relationships to a variety of other behaviors, at least for boys.

The studies just summarized reflect some of the growing interest in the role of small details in the perception and conceptual organization of familiar objects and pictorial stimuli. As already

indicated, the present study was concerned with exploring the role of small details as compared with that of over-all form in comparative similarity judgements made by children when dealing with simple geometric stimuli. More specifically, interest was centered upon the following problems: a) a comparison of the relative salience of over-all shape and small detail as stimulus cues; b) developmental changes in the utilization of form and detail as criteria for judging similarity over the age span from three to eight; and c) scope and correlates of individual differences in cue utilization.

METHOD

Subjects:

The sample employed in this investigation consisted of 153 children, approximately half boys and half girls, ranging in age from 2 years, 11 months to 8 years, 3 months. For purposes of analysis, they were divided into five age-grade groups: three-year-old and four-year-old nursery school children, along with kindergarten, first, and second grade groups from a public elementary school, all from the city of Ithaca, N. Y. The subjects represented an essentially middle class sample, and the schools were so selected that all age groups were approximately equivalent in social class status. Table 1 indicates the number of boys and girls in each of the five age-grade groups, along with the age range and mean ages for each group. Figures are given both for the total sample of 153 cases, as well as for a reduced sample of 125 subjects which was employed for analyses of variance requiring at least proportional numbers of boys and girls at each age.

An additional group of subjects from the same elementary school was employed in a replication study about a year after the original data on the large sample had been collected. This group consisted of 20 boys and 20 girls tested in the summer following completion of first grade. Included in this replication sample were 14 children who had previously been tested during their kindergarten year as part of the large sample.

Table 1: Composition of total and reduced samples in terms of school grade, age and sex

School Grade	Total Sample			Reduced Sample		
	N	Mean Age (mos.)	Age Range (mos.)	N	Mean Age (mos.)	Age Range (mos.)
Nursery 3 yrs. B	14	42.8		12	43.6	
G	<u>11</u>	42.3		<u>13</u>	43.2	
Total	25		35-47	25		35-49
Nursery 4 yrs. B	15	54.7		12	56.5	
G	<u>20</u>	56.3		<u>13</u>	57.1	
Total	35		48-65	25		50-63
Kindergarten B	16	69.7		12	68.6	
G	<u>17</u>	70.1		<u>13</u>	69.4	
Total	33		65-76	25		65-74
1st Grade B	15	81.5		12	80.9	
G	<u>15</u>	80.3		<u>13</u>	81.4	
Total	30		73-88	25		76-87
2nd Grade B	15	92.3		12	92.4	
G	<u>15</u>	92.6		<u>13</u>	92.8	
Total	30		85-99	25		88-98

PROCEDURE

a. Tasks for assessing response to form and detail.

In order to explore the role played by various types of details and of over-all shape as similarity cues, a number of comparative similarity tasks were developed, employing a matching-with-standard technique. Each task consisted of a row of four simple geometric figures printed on a 5 x 8 card, below which was a "standard" figure. Figure 1 illustrates the six tasks employed in the present study. The subject is instructed to look at all the drawings carefully, and then to indicate which one of the four figures he regards as most like the standard. When he has made his choice, that particular figure is covered with a small card, and he is asked which one of the remaining three figures is most like the standard. After each judgement, the child is asked to indicate briefly how the figure he has selected and the standard are alike.

It will be apparent from the illustrations in Figure 1 that in each task, the stimuli were so constructed that the child could make his first choice on the basis of a strong similarity in gross form (F-1 figure), but this would require that he ignore differences in detail characteristics of the standard and the F-1 stimulus; or, he could make a choice based on a clear similarity in detail properties (D-1), ignoring rather gross form differences between the standard and the D-1 stimulus. The remaining two choice figures were constructed so as to provide a "weaker" form similarity (F-2) than the F-1 stimulus, and a weaker detail similarity (D-2) than that represented in the D-1 stimulus. This was done in the interest

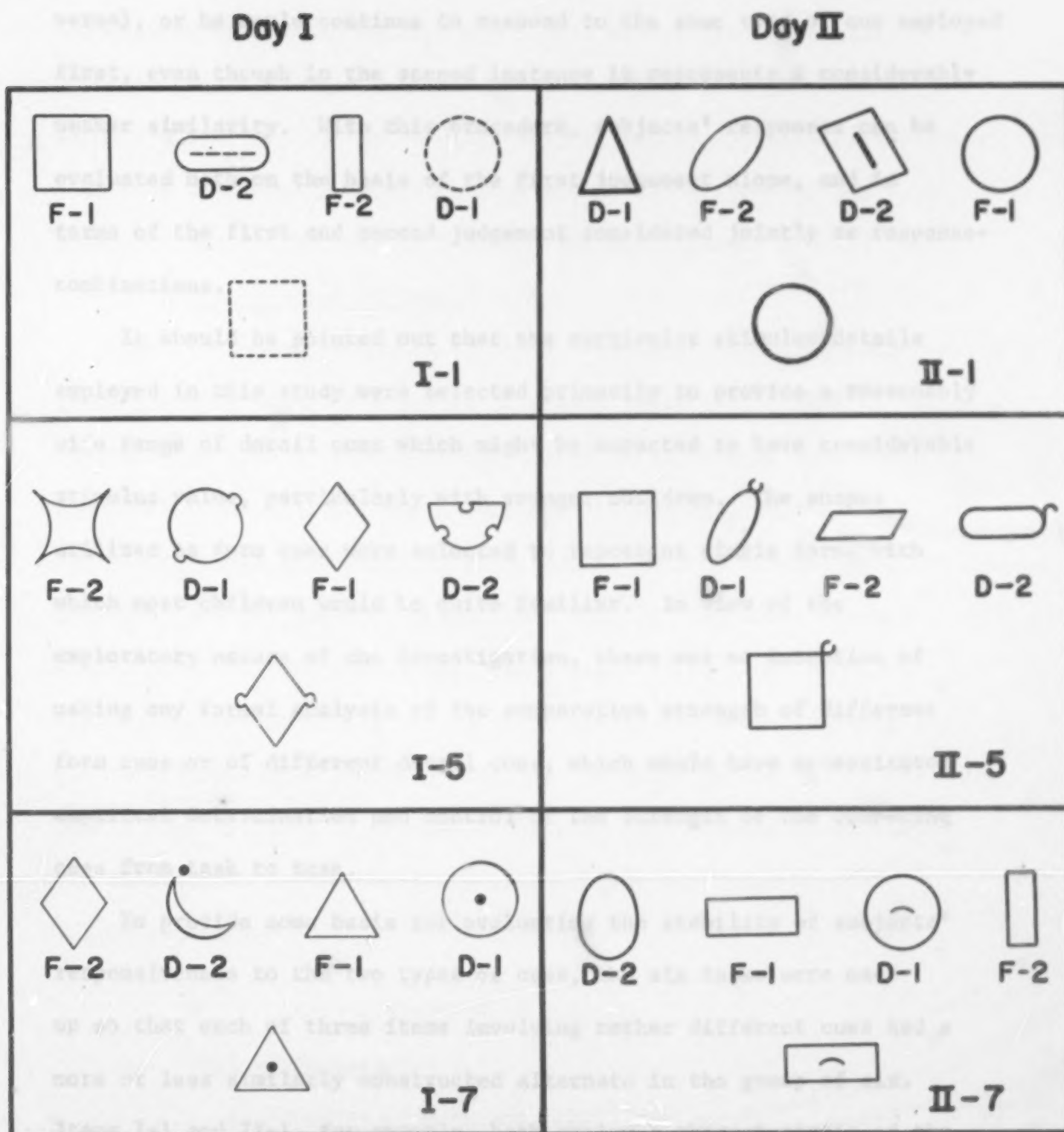


Figure 1: Geometric Items Used for Assessing Utilization of Gross Form or Detail as Determinants of Comparative Similarity Judgments

of "testing the limits" of responsiveness to either type of cue, since a subjects' second judgement could then be based upon a shift from one type of similarity cue to another (e.g., form to detail or vice versa), or he could continue to respond to the same type of cue employed first, even though in the second instance it represents a considerably weaker similarity. With this procedure, subjects' responses can be evaluated both on the basis of the first judgement alone, and in terms of the first and second judgement considered jointly as response-combinations.

It should be pointed out that the particular stimulus details employed in this study were selected primarily to provide a reasonably wide range of detail cues which might be expected to have considerable stimulus value, particularly with younger children. The shapes utilized as form cues were selected to represent simple forms with which most children would be quite familiar. In view of the exploratory nature of the investigation, there was no intention of making any formal analysis of the comparative strength of different form cues or of different detail cues, which would have necessitated empirical determination and control of the strength of the competing cues from task to task.

To provide some basis for evaluating the stability of subjects' responsiveness to the two types of cues, the six tasks were made up so that each of three items involving rather different cues had a more or less similarly constructed alternate in the group of six. Items I-1 and II-1, for example, both employ a characteristic of the line forming the contour of the figure as the detail similarity (the dotted line in I-1, the line made up of tiny circles in II-1).

In the next two items, the detail similarity cue is represented by small protrusions on the periphery or outline of the figure, which do not alter significantly the gross shape of the stimulus. It was anticipated that these details might have considerable stimulus value in terms of their connotative or associative meaning (bumps, ears, feelers, etc.). In the last pair of items, the detail similarity consisted of a small element located within the outline of the figure: a dot in the case of I-7, and the small curved line in II-7.

b. Administration of tasks.

The six geometric similarity tasks just described were administered individually as part of a series of twenty tasks in two testing sessions several days apart. (The remaining tasks in the series involved a number of two-choice geometric similarity items and eight comparative similarity tasks with familiar pictorial stimuli as content). Items I-1, I-5, and I-7 appeared as the 1st, 5th, and 7th tasks in the series on the first testing day, while Items II-1, II-5, and II-7 appeared in corresponding serial positions on the second testing day. The Ammons Picture Vocabulary Test (2) was given at the end of the second days' session to provide a short measure of general ability. Two experienced female examiners administered the tests in the school settings.

ANALYSIS OF RESULTS

a. Total Score Analyses.

1. Reliability of form and detail responses.

One of the first questions which arose in regard to children's utilization of shape or detail as similarity cues, was that

was that of the generality of response across tasks. When an analysis was made of the consistency of form and detail responses on the three items presented on day 1 and those given on day 2, a reasonably high degree of stability was found.⁴ Consequently it was decided to obtain over-all measures of responsiveness to form or detail based on all six items. Two such measures were developed: one indicating the number of tasks for which the first similarity choice made by the subject was based on form cues (i.e., number of F-1 choices as first response), the other based on an arbitrary weighting of both first and second responses, with high scores reflecting more frequent utilization of form rather than detail cues. Weights assigned in computing this total "FD" score were as follows: a) both responses based on form similarity = 3, b) first response based on form, second response on detail = 2, and c) first response based on detail, second response either form or detail = 1. The possible range of total FD scores was thus 6 to 18, that of the F-1 scores 1 to 6.

Most of the total score analyses were based on the FD score because of its more widely differentiated range (see Table 3), and because it was less highly skewed than the F-1 score, particularly at the upper age levels. That the FD score provided a satisfactorily reliable measure of subjects' tendency to respond to form rather than to detail is indicated by the estimated split-half reliability

⁴For each of the three pairs of items, Chi-squares for the total sample were all significant at the .01 level or below, and the three contingency coefficients were .48, .46, and .43.

of .80, based upon the correlation between the day 1 and day 2 items, for the entire sample of 153 subjects. (The reliability tends to be lower than .80 in the three younger groups, and higher in the two older groups). Suggestive evidence of a relatively high degree of long term stability in the child's use of form or detail cues was found in the test-retest correlation of .53 for the 14 children tested as kindergarteners in the large sample, and tested again over a year later as part of the small replication sample.

2. Age changes and sex differences.

The next phase of the analysis focused upon an evaluation of the over-all salience of stimulus shape and of small detail as competing similarity cues, with particular reference to changes in cue utilization as a function of age and sex. Several types of data pertinent to this problem are presented in the paragraphs which follow, and in the accompanying figures. The mean F-1 scores for boys and girls in each age-grade group were computed and plotted for both the total and reduced samples (Figure 2), and a similar analysis was carried out for the FD scores (Figure 3). Turning first to Figure 2, which is based upon the number of tasks for which the first response is a form similarity, it is immediately apparent that the age curves for boys and girls are appreciably different. For girls, stimulus shape rather than detail was the more frequent determinant of first similarity judgements, with little or no variation throughout the age range studied. On the average, girls made their first choice on the basis of form on four or five of the six tasks presented. Boys, on the other hand, were considerably less responsive to form cues than the girls were in the early

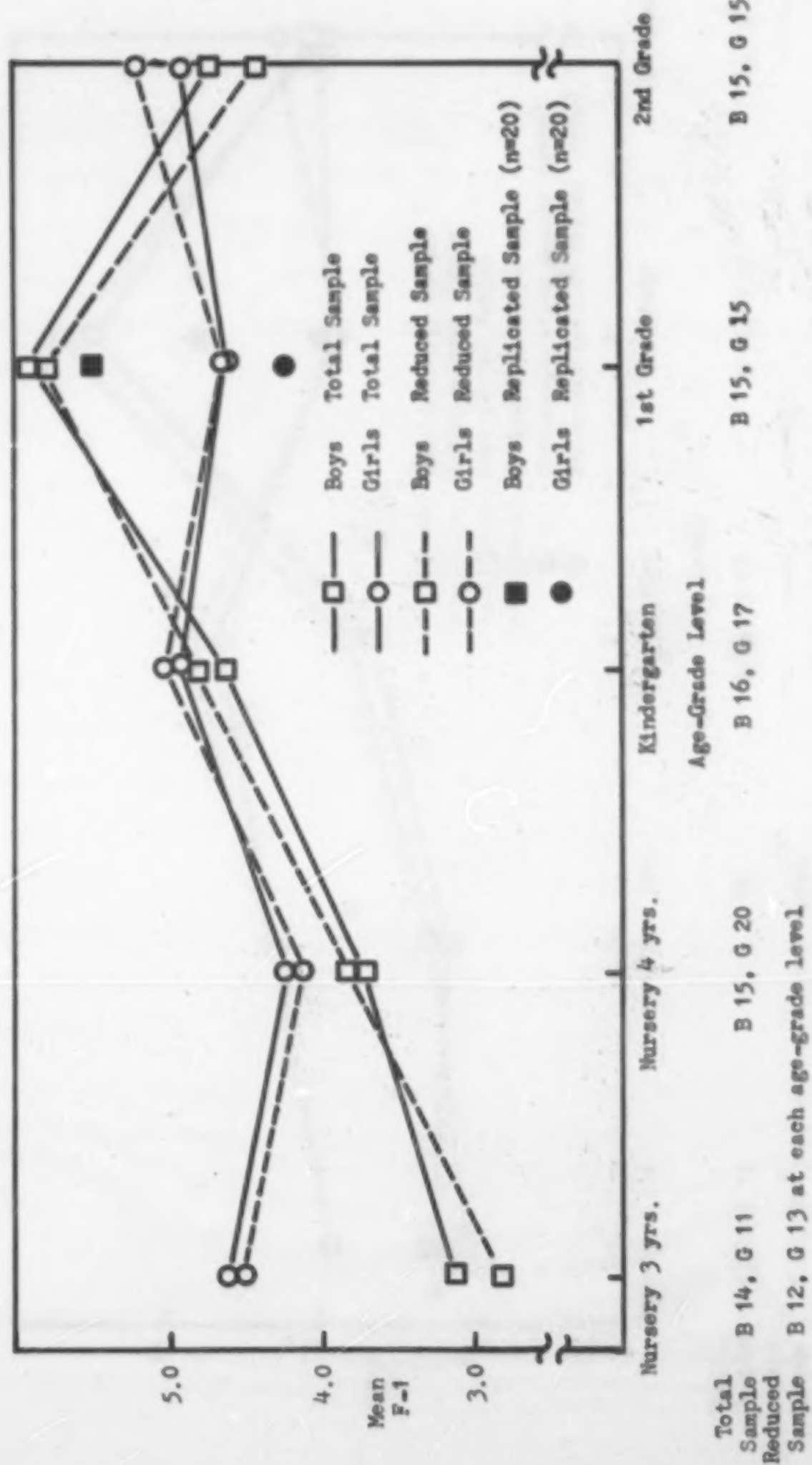


Figure 2: Mean F-1 Scores for Boys and Girls At Successive Age-Grade Levels
Total and Reduced Samples

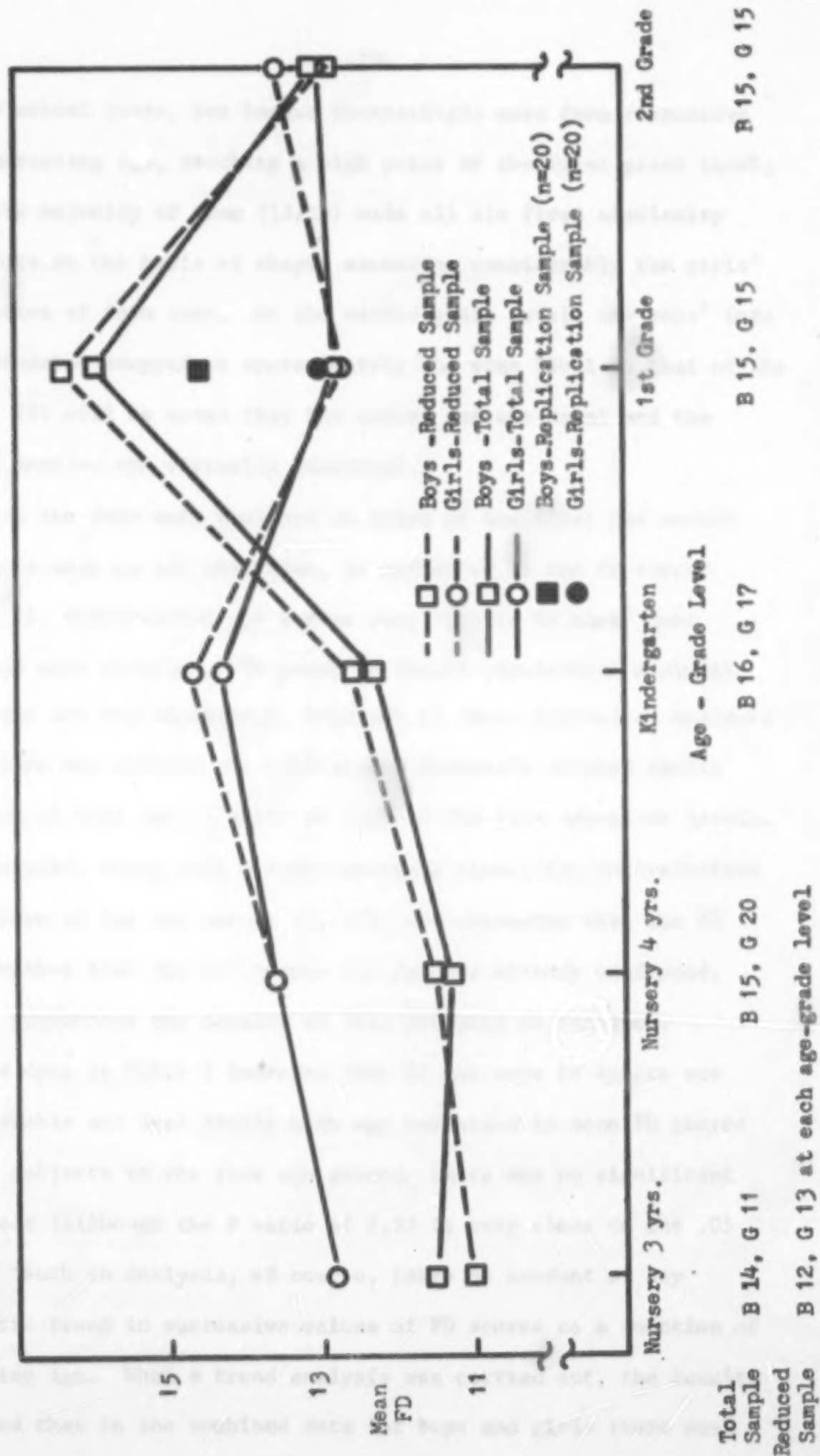


Figure 3: Mean FD Scores for Boys and Girls At Successive Age-Grade Levels
Total and Reduced Samples

nursery school years, but became increasingly more form responsive with increasing age, reaching a high point at the first grade level, where the majority of them (13/15) made all six first similarity judgements on the basis of shape, exceeding considerably the girls' utilization of form cues. At the second grade level, the boys' form responsiveness dropped to approximately the same level as that of the girls. (It will be noted that the curves for the total and the reduced samples are virtually identical.)

When the data were analyzed in terms of the first and second judgements made on all six tasks, as reflected in the FD scores (Figure 2), differential age curves very similar to those just described were obtained. To permit a formal statistical analysis of the age and sex differences apparent in these curves, an analysis of variance was carried out utilizing a minimally reduced sample providing 12 boys and 13 girls at each of the five age-grade levels. This analysis, along with a trend analysis permitting an evaluation of the form of the two curves (1, 12), was conducted with the FD scores rather than the F-1 scores for reasons already mentioned. Table 2 summarizes the results of this analysis of variance.

The data in Table 2 indicate that if one were to ignore sex as a variable and deal simply with age variation in mean FD scores for all subjects in the five age groups, there was no significant age effect (although the F ratio of 2.33 is very close to the .05 level). Such an analysis, of course, takes no account of any systematic trend in successive values of FD scores as a function of increasing age. When a trend analysis was carried out, the results indicated that in the combined data for boys and girls there was

Table 2: Summary of Analysis of Variance and Trend Analysis,
FD Scores, Reduced Sample

Source of Variation	df	Sum of Squares	Mean Square	F
Total Group				
Between Groups	9	261.64	29.07	2.80**
Within Groups	115	1193.75	10.38	
Total	124	1455.39		
Age Groups	4	96.83	24.21	2.33
Sex	1	10.98	10.98	--
Sex X Age	$\frac{4}{9}$	153.83	38.46	3.71**
Age Groups Breakdown				
Linear	1	52.90	52.90	5.10*
Quadratic	1	29.15	29.15	2.81
Cubic	1	14.40	14.40	--
Quartic	$\frac{1}{4}$	.39	.39	--
Boys Only				
Between Groups	4	216.44	54.11	5.20**
Within Groups	55	571.75	10.40	
Total	59	788.19		
Between Groups Breakdown				
Linear	1	91.88	91.88	8.83**
Quadratic	1	23.63	23.63	2.28
Cubic	1	75.21	75.21	7.23**
Quartic	$\frac{1}{4}$	25.73	25.73	2.47
Girls Only				
Between Groups	4	34.22	8.56	--
Within Groups	60	622.00	10.37	
Total	64	656.22		
Between Groups Breakdown				
Linear	1	.77	.77	--
Quadratic	1	7.93	7.93	--
Cubic	1	9.42	9.42	--
Quartic	$\frac{1}{4}$	16.09	16.09	1.55

** F ratio significant at .01 level

* F ratio significant at .05 level

This is indicated by the significant cubic component F ratio of 7.23. The quadratic component is insignificant except the U shaped characteristic occurs only at the extreme end of the curve rather than being a property of the entire curve.

a significant linear age trend, or a significant linear relationship between FD scores and age ($F=5.10$, $P<.05$). (Higher order polynomial trends were not significant for the combined data). It is obvious from the curves in Figure 3, however, that the increase in FD score with age is attributable almost entirely to the boys. This difference in the effect of age upon the form responsiveness of boys and girls was highly significant (as indicated by the sex x age interaction in Table 2).

Trend analyses were also carried out separately for boys and girls, with the results summarized again in Table 2. There was a significant main age effect for boys, and not for girls, as expected on the basis of the analysis just discussed. Of greater interest was the finding of a highly significant linear trend for boys ($p<.01$), while neither the linear nor any of the remaining trends were significant in the girls' data. In the case of the boys' curve, the extreme rise in form responsiveness found at the first grade, followed by the sharp drop at the second grade, is a particularly puzzling finding. The trend analysis suggests that this inverted U shaped portion of the curve is not random deviation from linearity, but represents a significant curvilinearity superimposed upon a basically linear trend, at the upper end of the curve.⁵ Some post-hoc attempts to explain this curvilinear trend will be presented later in the paper. It should be added at this point that the sex difference

⁵This is indicated by the significant cubic component F ratio of 7.23. The quadratic component is insignificant since the U shaped characteristic occurs only at the extreme end of the curve rather than being a property of the entire curve.

in FD scores at the first grade level was significant ($p < .01$), while those found in the three youngest groups did not reach this level of significance although girls were consistently more form responsive than boys (N-3's, $P < .10$, N-4's, $P < .20$, and kindergarteners, $P < .10$).

When the data just discussed were first examined, the marked peak in form responsiveness for boys at the first grade level, which represented a sharp reversal of the trends in the younger groups and in the second grade, was regarded as a probable chance result. To determine whether this sex difference in favor of boys in the first grade was a reliable finding, another sample of 20 boys and 20 girls from the same school were tested in the summer following their completion of first grade. (Actually, these children did not represent a completely independent sample, since, as already mentioned, 5 boys and 9 girls had been tested as part of the original sample over a year previously.) It was of considerable interest to find that in this first grade replication sample, boys were once again considerably more responsive to form cues than was the case for girls ($P < .10$), confirming the large sample results. In fact, as indicated in Figures 2 and 3, the mean F-1 and FD values for the replication sample were quite close to those obtained in the original sample, although generally somewhat lower.

Since 14 children in the replication sample had previously been tested in kindergarten, it was decided to examine changes in the form responsiveness of these subjects to determine whether there was any longitudinal confirmation of the differential age changes for boys and girls indicated in the large sample cross sectional data. From these cross sectional age curves, one would predict that from

kindergarten to first grade, the boys ought to become more responsive to form, while the girls ought to show essentially no change, or perhaps a slight drop. This is precisely what was found for both the F-1 and FD scores. (In the former case, boys changed from 3.8 to 5.4, girls from 4.6 to 3.6; for the FD scores, boys changed from 11.4 to 14.0, girls from 12.3 to 11.9.) Although these findings are based upon a very small number of children, they do provide some longitudinal support for trends observed in the cross sectional data.

It thus appears, in summary, that the greater utilization of form cues by first grade boys than by first grade girls was a reliable finding. The significance of the curvilinear component in the boys' age curve in Figure 3, along with the longitudinally obtained empirical evidence of a rise in FD score from kindergarten to first grade in the case of boys only, suggest that this rather sudden, large increase in form responsiveness at the first grade, followed by a decline at the second grade, is a genuine rather than a chance phenomenon. Ideal confirmation of this curvilinear trend, of course, would require replication of the curves over at least the three upper age levels. It should be mentioned, finally, that the obtained sex differences are not explainable in terms of age or ability differences since boys and girls were essentially equivalent on these two variables in every age-grade group.

b. Analyses of four response alternatives.

To supplement the data just presented on age changes and sex differences revealed by analyses of the F-1 and FD total scores,

an analysis was made of the frequencies with which each of four choice stimuli (F-1, F-2, D-1, D-2) were selected by boys and girls as first responses at each age level. These data, in the form of percentages of judgements made on all six tasks combined are presented in Figures 4 and 5 for boys and girls respectively. The curves indicating the percentage of F-1 responses at each age level will be recognized as representing essentially the same trends previously shown by the mean F-1 scores in Figure 2. The strong predominance of gross form as a similarity cue at all ages and for both boys and girls is very clearly revealed in Figures 4 and 5. The percentage of F-1 responses varied from 70 to 82 percent across age groups for girls, and from 52 to 97 percent for boys.

The next most frequent first response, at all ages and for boys and girls, was the D-1 stimulus, which resembled the standard in a detail property but differed from it in over-all shape. The remaining two stimuli, F-2, and D-2, occurred very infrequently as first responses, primarily in the youngest groups. In the case of the girls, about 17 percent of their first similarity judgements were determined by detail cues at every age level. For the boys, a slightly higher proportion (about 20%) of their first responses were determined by detail cues at all ages except for the first grade level where, as previously noted, virtually all responses (97%) were form-determined. In the younger groups, although the differences were small, boys responded consistently more often than girls to both the D-1 and D-2 choices, again reflecting their greater sensitivity to detail cues in the early years. In summary, then, it is clear that for the particular stimuli used in the present study, form cues were very

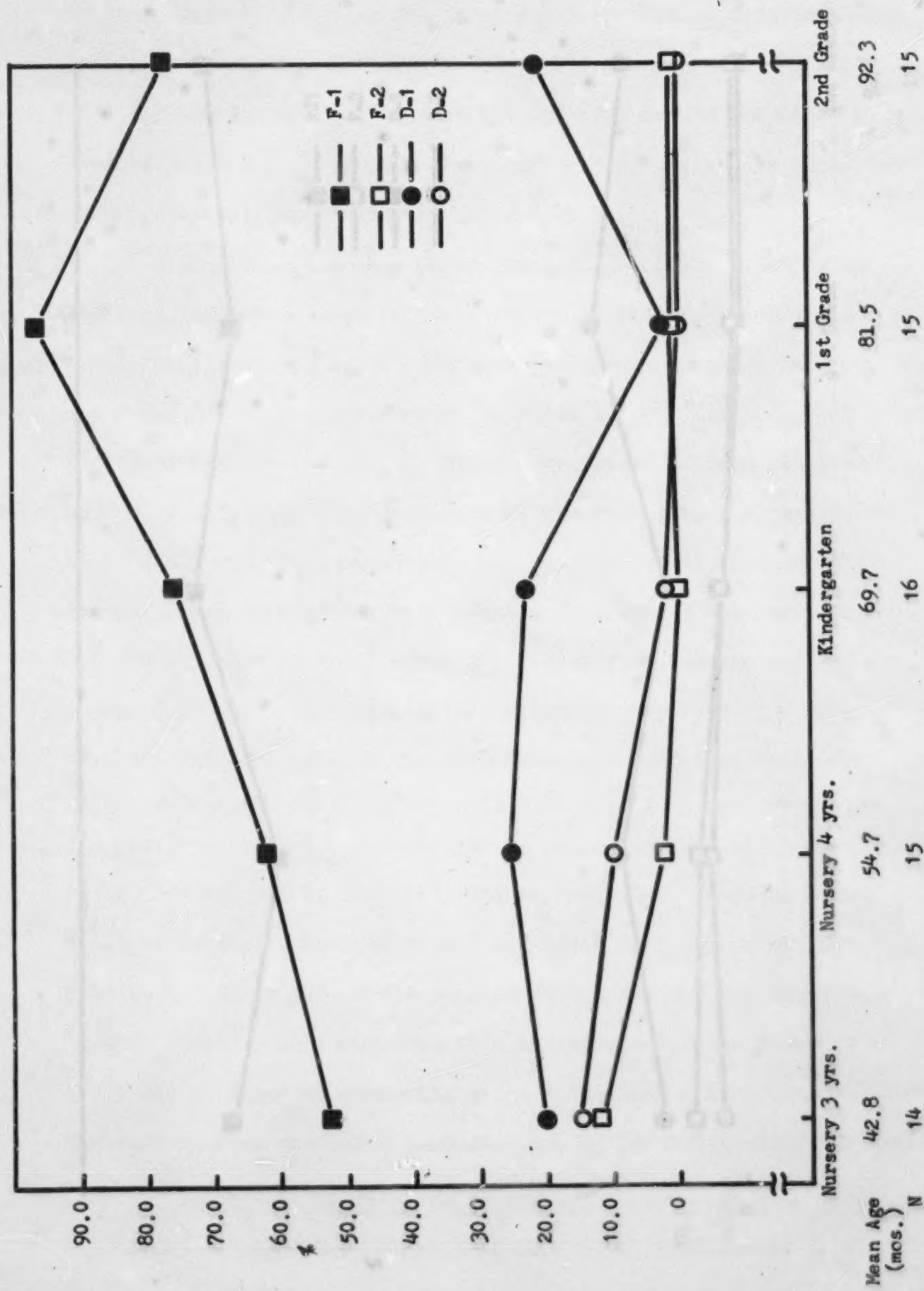


Figure 4: Proportion of F-1, F-2, D-1 and D-2 Responses Given as First Choice for All Six Items
Total Sample, Boys Only

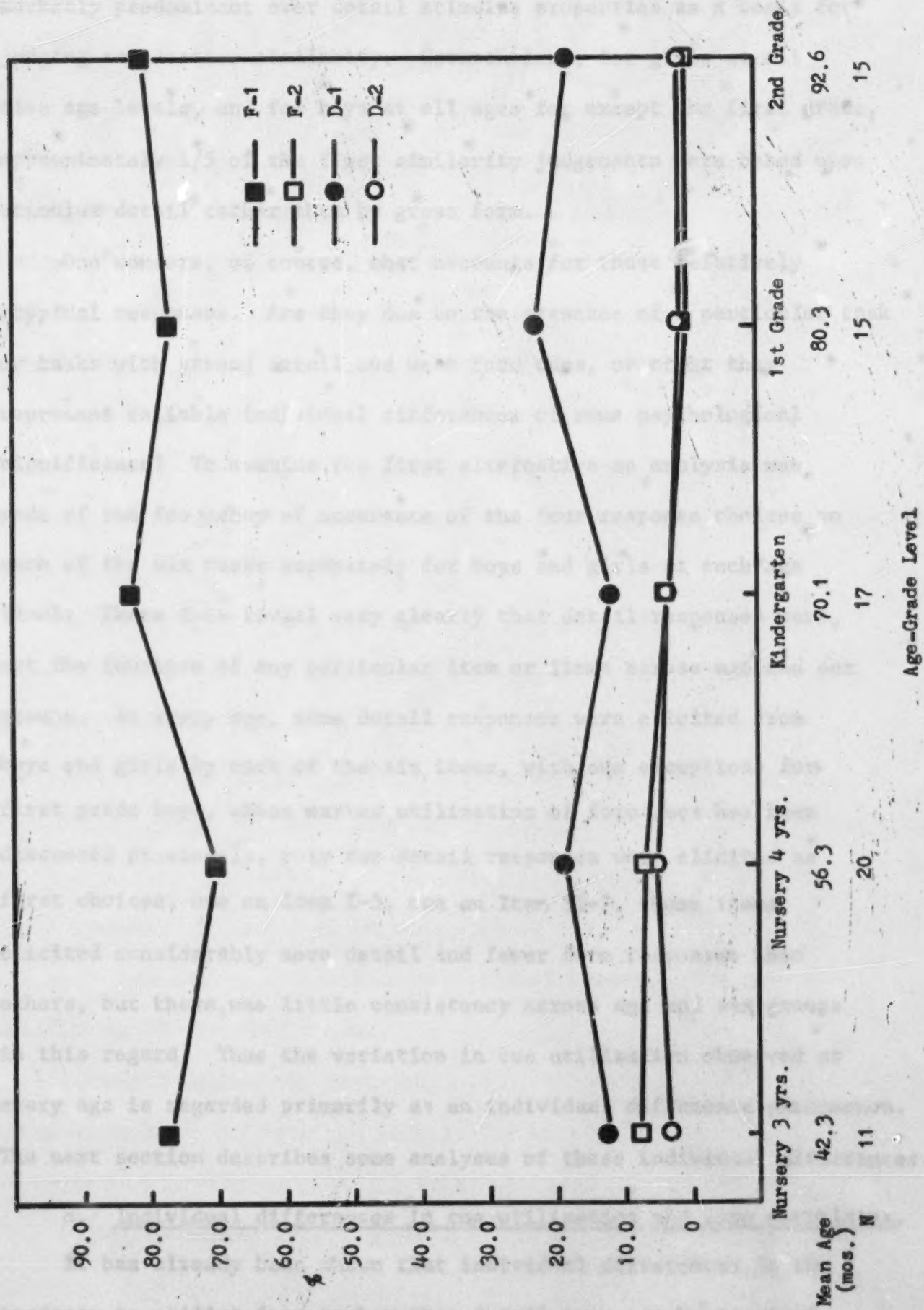


Figure 5: Proportion of F-1, F-2, D-1 and D-2 Responses Given as First Choice for All Six Items
Total Sample, Girls Only

markedly predominant over detail stimulus properties as a basis for judging comparative similarity. Nevertheless, for girls at all five age levels, and for boys at all ages for except the first grade, approximately 1/5 of the first similarity judgements were based upon stimulus detail rather than by gross form.

One wonders, of course, what accounts for these relatively atypical responses. Are they due to the presence of a particular task or tasks with strong detail and weak form cues, or might they represent reliable individual differences of some psychological significance? To examine the first alternative an analysis was made of the frequency of occurrence of the four response choices on each of the six tasks separately for boys and girls at each age level. These data reveal very clearly that detail responses were not the function of any particular item or items across age and sex groups. At every age, some detail responses were elicited from boys and girls by each of the six items, with one exception: for first grade boys, whose marked utilization of form cues has been discussed previously, only two detail responses were elicited as first choices, one on Item I-5, one on Item II-7. Some items elicited considerably more detail and fewer form responses than others, but there was little consistency across age and sex groups in this regard. Thus the variation in cue utilization observed at every age is regarded primarily as an individual difference phenomenon. The next section describes some analyses of these individual differences.

c. Individual differences in cue utilization and some correlates.

It has already been shown that individual differences in the tendency to utilize form rather than detail cues can be measured

with relatively good reliability. We are then faced with a number of interesting questions concerning the explanation of these individual differences. Why is it that at every age there are some children who are very highly responsive to form cues in making similarity judgements, while others make considerably less use of such cues or are primarily responsive to details in the stimulus? Can we identify some of the determinants of these individual differences in terms of such factors as differential experience, or differential response to the "average predictable environment", for example? What are some of the consequences of these individual differences in terms of relationships to other variables of psychological importance?

In the present investigation, only a preliminary attempt to begin to answer some of these questions was possible, by examining the relationships of individual differences in form responsiveness to several other variables. These variables were: general ability, perceptual acuity, discrimination of three-letter nonsense syllables, and similarity judgements on a pictorial task. Before describing these analyses, however, it will be well to indicate very briefly the scope and nature of the inter-individual variation found in utilization of form and detail.

Table 3 presents the distributions of FD scores for boys and girls at each of the five age levels. It will be noted that the range of variation is very wide at all ages, although the distributions tend to become skewed at the upper age levels, particularly for boys.

Table 3. Distribution of FD Scores for Boys and Girls
at Each Age-Grade Level

Total Sample										
	Nursery 3 yrs.		Nursery 4 yrs.		Kindergarten		1st Grade		2nd Grade	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
18			1	3		5	7	2	4	1
17		1	1	2	1	1	1	3		1
16	1			3	2	1	3		1	3
15	1	1		1		1		1		
14	2	3	1	3	1	2			2	2
13	1	3	2		5	3	2	1	2	2
12	1		2	1	1		1	1	1	1
11	3	1	1	3	2	2	1	2	2	1
10	2	2	3	1	2			1		
9	1		1	2				1		3
8	1		1		1	2		3	1	1
7	1				1				1	
6			2	1					1	
Total	14	11	15	20	16	17	15	15	15	15

Since most children utilize form more than detail as similarity cues, much of the variation in FD scores actually represents individual differences in the degree to which form is responded to as the dominant cue. Nevertheless, at every age there are some children for whom the predominant cues are the stimulus details rather than over-all shape (FD scores below 9 or 10, approximately).

1. Relation to general ability.

One of the first questions which arose in attempting to understand the significance of these individual differences was that of its possible relationship to general ability level. As previously indicated, the Ammons Picture Vocabulary Test (2), with very slight modifications in administration and scoring, was given to all children in the study. This instrument, although short, has reasonably high correlations with standard individual intelligence test measures. In the present investigation, raw scores on the Ammons were employed as approximate measures of general intellectual level throughout the entire age range. When FD scores were correlated against the Ammons measures of general ability at each of the five age levels, the results indicated small positive relationships in the two nursery school groups and at the first grade, with essentially zero relationships at the other two age levels (Pearson r 's from the youngest to the oldest group were .35, .33, -.05, .26, and .05). When the relationship between FD scores and general ability was analyzed for the entire sample, the correlations were .27 for boys, .13 for girls, and .19 for the combined group.

A more precise evaluation of these relationships requires, however, that the effect of age be partialled out. Table 4 presents the various required zero order correlations, along with the partials for boys, girls, and the total group. The previously discussed analysis of variance indicated a significant linear relationship between FD scores and age for boys, but not for girls. The magnitude of this relationship is indicated in Table 4 by the Pearson r of .29.

Table 4. Intercorrelations Among FD Scores (Form Responsiveness),
Ammons Picture Vocabulary, PST (Picture Similarity Test), and Age
Total Sample

		2 Ammons	3 PST	4 Age	Partial
Boys (N=75)*	1 FD	.27	.34	.29	$r_{12.4} = .04$
	2 Ammons		.53	.86	$r_{13.4} = .23$
	3 PST			.54	
Girls (N=78)*	1 FD	.13	.25	-.02	$r_{12.4} = .25$
	2 Ammons		.45	.81	$r_{13.4} = .29$
	3 PST			.43	
Total (N=153)**	1 FD	.19	.29	.14	$r_{12.4} = .13$
	2 Ammons		.49	.83	$r_{13.4} = .26$
	3 PST			.48	

* $r = .29$ for $p .01$; $r = .22$ for $p .05$.

** $r = .21$ for $p .01$; $r = .16$ for $p .05$.

As expected, the Ammons general ability measures were highly correlated with age (r 's = .86, .81, and .83). When the effect of age was partialled out, the correlations between FD scores and general ability were .04 for boys, .25 for girls, and .13 for the total group. It thus appears that individual differences in form or detail cue utilization are not accountable for in terms of differences in general intelligence.

2. Relation to perceptual acuity.

Although the various details provided as potential similarity cues were all clearly distinguishable, the question was raised as to whether those children who tended to utilize detail cues rather than over-all form as determinants of similarity judgements were particularly skillful in making relatively fine perceptual differentiations. To answer this question, a special test was developed to measure individual differences in children's ability to make fine perceptual discriminations between relatively simple geometric stimuli. This "perceptual acuity" task consisted of 22 items, each involving the presentation of three simple geometric figures drawn on a 5 x 8 card, one of which differed very slightly from the other two (e.g. three parallel lines, one being $1/8$ inch longer than the other two; three circles, one of which had a diameter $1/16$ inch shorter than the others). For each item, the subject was asked to judge which stimulus was not like the other two, and the total score was based on the number correct. It was possible to administer the perceptual acuity test only to the small replication sample of 20 boys and 20 girls. For each sex group the scores ranged from 12 to 22, with reasonably normal distributions having means of 17.7 and 17.1 for the boys and girls respectively.

When the relationship between FD scores and perceptual acuity was analyzed, the correlations were $-.11$ for boys, $.18$ for girls, and $.03$ for the entire sample of 40 cases. It thus seems pretty clear that the tendency to utilize detail as a similarity cue, or to be relatively unresponsive to form, was not explainable in terms of

a heightened capacity to make fine perceptual discriminations, at least at the age level represented by the small replication sample.

3. Relation to discrimination of nonsense syllables.

Thus far, the data has shown that responsiveness to form or detail as similarity cues increases significantly with age, for boys only, and is virtually unrelated to either general ability or to perceptual acuity. The next question which arose was whether the tendency to be particularly responsive to shape or to detail cues in making similarity judgments had any relationship to the child's capacity to make the kinds of discriminations involved in learning to read. Although this was not a major purpose of the study reported here, it was possible to examine the question in a limited fashion as part of a thesis by Woodward (19) who was concerned with the effect of the number and types of differentiating cues on discrimination of three-letter nonsense words by the kindergarten and first grade children employed in the present research. When Woodward examined the relationship between tendency to utilize form rather than detail cues and over-all accuracy in word discrimination, she found significant positive correlations of appreciable magnitude between these two variables for her combined boys' group (r 's = .47 for 1/50th second word exposure, .66 for 1 second word exposure, $N = 27$). The corresponding correlations for girls were insignificant and close to zero. In the case of the boys, the correlation between form responsiveness and accuracy was accounted for primarily by the kindergarten group, with the relationship being much lower among the first graders, perhaps because of the highly skewed distribution and restricted range

of FD scores for first grade boys.

When the relationship between tendency to utilize form or detail cues and types of errors made in word discrimination was analyzed, some very provocative findings were obtained by Woodward. All subjects found word discrimination easiest when the differentiating letter was the first letter in the word, and considerably more difficult when the differentiating letter or letters were in the second and/or third position in the word. However, the results obtained particularly in the kindergarten group, suggest that when the differentiating cue is in the first position, children who are highly form responsive are considerably more accurate than those who are less responsive to form. This relationship was consistent for both boys and girls at the kindergarten level, but for first graders these discriminations were so easy that both high and low FD subjects made practically no errors. When the discriminating cues were in the second and/or third positions, in three out of four comparisons for kindergarten subjects, the trends just described were reversed and it was the less form responsive or "detail prone" children who made more accurate discriminations. Similarly for first grade girls but not boys, the less form responsive subjects were more accurate when the differentiating letters were in the second or third position.

Although Woodward's findings must be considered highly tentative because of the small number of subjects involved, they clearly suggest that children who are particularly responsive to form cues in making comparative similarity judgements may be generally more accurate in making word discriminations, particularly when the

distinctive cue is in the easily perceived first position. However the more detail responsive subjects appeared to make better use of differentiating cues when they were in the less easily perceived second and/or third positions in the word. This suggests, in at least a speculative way, that one of the implications of responding to the relatively atypical, detail cues in judging similarity of the geometric stimuli is a tendency to search out and utilize the less obvious, less easily perceived properties of stimuli, with consequently more successful discrimination on tasks where such properties represent the differentiating features. Further studies to examine the validity of this hypothesis would certainly seem indicated.

4. Relation to pictorial similarity judgements.

The next inquiry was aimed at determining whether there was any generality in the mode of stimulus cue utilization across similarity tasks with quite different content. The children in this study had also been given a series of pictorial similarity tasks, on each of which the subject had to make a first and second choice to indicate which of several stimulus pictures were judged to be most like a standard picture. The total score (PST) to be discussed here, based on four tasks, reflected the extent to which the subject judged similarity on the basis of conventional, representational meaning rather on the basis of such "perceptual" cues as color, visual "textural" qualities, irrelevant details, etc. When the FD measures were correlated with the pictorial similarity scores at each age level, there was some evidence for a relatively low, generally positive relationship. (Pearson r 's from the youngest to the oldest

groups were as follows: .16, .45, .37, .23, and -.13.) Combining all age groups, the correlations were .34 for boys, .25 for girls, and .29 for the entire sample, as indicated in Table 4.

Since PST scores were correlated with age to a considerable degree, the relationship between form responsiveness and the tendency to judge pictorial similarities on the basis of conventional meaning rather than on stimulus details had to be evaluated with the affect of age partialled out. The partial correlations between FD and PST scores thus obtained were .23 for boys, .29 for girls, and .26 for the total group. It is therefore apparent that even when the common co-variation due to age is partialled out, there is a small but significant relationship between the tendency to utilize over-all shape rather than detail cues in judging similarity of geometric figures, and the tendency to respond to the conventional meaning of familiar pictures rather than to "fragmentary" stimulus properties such as color, small details, etc., when making similarity judgements on familiar pictorial stimuli. Thus, in addition to relatively high stability across different groups of geometric stimuli, individual differences in preferred mode of cue utilization appeared to have at least some generality across both geometric and familiar pictorial material.

SOME QUALITATIVE ASPECTS OF THE JUDGMENTAL PROCESS

The nature of the comparative similarity judgement itself, and the processes or mechanisms which it involves, remain elusive to us. One of the basic problems is that of determining exactly what stimulus properties or combination of properties the child is

responding to when he says that two objects are alike, or are different. While it is possible to shed some light on this question by comparing judgements of stimuli whose properties are varied and controlled by the experimenter, as in the case of the present research and many others, one always wonders whether the child is in fact responding to the particular stimulus characteristic which the experimenter has manipulated.

In hope of exploring this question a bit further, children in the present study were asked after each similarity judgement to indicate briefly how the chosen figure was like the standard. There are obvious difficulties involved in this approach, since it is entirely conceivable, and sometimes clearly evident, that the child's post-hoc verbal explanation of his judgement may make reference to cues which did not in fact determine the original similarity judgement. Consequently, such verbalizations may provide only ambiguous evidence as to factors which influenced the similarity judgement itself. Nevertheless, they offer at least suggestive evidence in this regard, and very good data on the range of characteristics which children report as constituting similarities, as well as the language with which these are expressed.

A detailed analysis of the children's verbalizations is beyond the scope of the present paper. However, it would be useful to indicate briefly several findings revealed by these verbal explanations. When children made detail choices, they gave ample evidence that they were in fact responding to the detail properties built into the stimuli by reference to them in their verbal explanations.

Reference to these cues sometimes involved simply pointing to the details and identifying them non-specifically (e.g. "they've both got those little things," "they both have these around it"); sometimes the details were described or identified more specifically in terms of obvious characteristics (e.g. "this has a dotted line around it too", "both got a curve in the middle", "this is made of circles and this is", "they both have half-round things"); and sometimes the details were identified or named as representing real objects in the child's world (e.g. "they both have chains", "they both got ears on them", "it has a little 'C' on the end"). Generally speaking, the last two types of verbal explanations occurred more often in the older children, with the non-specific identifications being found more typically in the younger groups, who also gave many explanations which were essentially indeterminate or non-informative.

With regard to form choices, it was again generally clear from the verbalizations that children were responding in various ways to similarities in over-all shape. It was also quite apparent that the processes involved in judging that two like-shaped figures are the same may differ considerably from child to child, or task to task. A large number of children, particularly in the older groups, identified the figures in terms of conventional geometric labels or adjectives ("both squares", "both triangles", "both round"), and some described them as representing familiar objects ("they're boxes", "two wheels for a car", "both like a kite" "a window and a window", etc.) Many children tried to indicate similarity in shape by simply describing the configuration or arrangement of parts ("it has corners and

straight lines", "like this all around (outlining)", "over and up the hill and down on both",) In some instances, the child referred to a similarity in one or several parts of the figure ("both got two sharp ends", "that's straight and that is", "this has points and this does too") without referring to the over-all configuration.

One of the questions raised by these variations in verbal response is the familiar problem of the role played by verbal mediation, or by the availability of verbal labels for categorizing stimuli in similarity judgements of the type investigated in this study. These responses also re-emphasize the importance of determining more precisely what it is about stimulus shape that the child is responding to when he judges two figures of similar form as "alike". One final observation about judgmental process suggested by the children's verbal responses concerns their simultaneous awareness of both similarities and differences in the pair of stimuli judged to be alike. Many children, for example, in making an F-1 choice, commented that if a detail weren't present, or if one were added, the figures would be identical. This "conditional similarity" response was particularly apparent in connection with the F-2 choice, when children often pointed out how the figures could be made more alike by appropriate transformations (e.g., "if this was turned it would be the same").

DISCUSSION

Insofar as the particular geometric stimuli and tasks employed in the present study are concerned, it is quite clear that over-all shape or outline of the figures was a far more potent determinant of similarity judgements than were the various detail characteristics.

This was true for both sexes and at all ages from 3 to 8. Obviously, one cannot derive any broad generalizations concerning the prepotence of form over detail from these findings, since the results obtained are so highly dependent on the particular stimuli employed and on the relative strength of the two types of competing stimulus properties built into the tasks. However, the fact that shape played as important a role as it did in the judgement of even the youngest children in the study is a finding of considerable interest.

In attempting to explain these results, two factors need to be considered. First, the shapes employed were for the most part simple, familiar forms to which most children have much exposure in every-day life, and which they learn to describe and identify with common verbal terms quite early. Had the forms been somewhat more complex, less familiar, or lower in "association value", they might have been considerably less potent as similarity cues. Secondly, with regard to the competing detail properties, they were not perceived by most children as representing "essential" features of the total stimulus, but rather as cues which were more or less irrelevant and subordinate to the shape. If these competing details had been constructed so that they could be regarded as representing important or "characteristic" details of the whole figure or of other well known objects (in Vurpillot's terms, 14), it is probable that the stimulus value of such "meaningful" details would considerably outweigh that of over-all shape alone. Further research in which the detail cues and the form cues are varied systematically along the dimensions just indicated and pitted against each other would certainly appear to be worth

pursuing. Considering the strong predominance of geometric form as a similarity cue revealed under the particular conditions of the present study, it would seem reasonable to hypothesize that these results reflect the strong emphasis placed on the over-all shape or form of things in our culture, and particularly the emphasis placed on the identification of simple, common geometric forms in the early life of the typical American child.

The differences found between the age curves for boys and girls in regard to form responsiveness represent a most interesting and puzzling phenomenon. Consider first the absence of any change in form responsiveness for girls over the age range studied, and the fact that boys made much less use of form cues than girls did in the early years, but showed a significant increase in form responsiveness with age. This greater utilization of form cues by girls in the younger groups was a rather unexpected finding, since one might have anticipated that sex differences, if found, would have been in the opposite direction. One possible explanation of these sex differences is offered by the hypothesis that girls respond earlier, and to a greater degree, to the emphasis which the culture places on the identification of simple geometric forms, hence they develop a selective responsiveness to form as a predominant similarity cue earlier than boys, whose sensitivity to form is still increasing during the age period studied.

Given this possible explanation of the differences in the age curves for boys and girls during the first three age periods, how then can we account for the sharp rise in form cue utilization by first grade boys, which is significantly greater than that of first grade girls,

and is then followed by a return to the same level as that of the girls at the second grade level? A plausible post-hoc explanation may be found in the hypothesis that this markedly heightened, although apparently temporary, selective utilization of form cues by first grade boys reflects their having been exposed to more intensive training in form discrimination and recognition than girls, as part of their pre-reading or reading experience in the first grade. It is conceivable, too, that boys react more strongly than girls to first grade training in form discrimination since it is being presented at a point in development where, for boys, increasing sensitivity to form is still growing, whereas girls have reached a stable level in this regard several years before the first grade. Informal discussion with school personnel following formulation of this hypothesis lent considerable credence to the idea that boys do in fact tend to receive somewhat more intensive reading readiness training than girls in the first grade, since teachers expect that learning to read will be more difficult for boys.

Our understanding of the possible psychological significance of individual differences in cue utilization is left still very incomplete, although the results do provide some helpful but tentative ideas in this regard. Preferential utilization of form or detail as criteria for judging geometric similarity seems to be unrelated either to general ability or to perceptual acuity. Degree of form responsiveness shows considerable stability across different geometric similarity items, and some degree of consistency with preferential utilization of conventional meaning rather than "perceptual"

stimulus qualities as criteria for judging pictorial similarities. There is some evidence, which needs to be confirmed, suggesting that children who are less responsive to over-all form and more sensitive to detail, make particularly effective use of relatively atypical, less obvious discriminating cues in word recognition, while those who are primarily form responsive make better use of the more obvious, conventional differentiating cues. (It would have been extremely interesting to relate form responsiveness to various measures of actual reading performance, but it was not possible to obtain satisfactory measures for this purpose.) Considered jointly, these findings suggest the possibility that individual differences in form and detail responsiveness may represent psychologically meaningful differences in children's selective response to and organization of stimulus cues in categorizing objects, with the form responsive children being more prone to utilize the culturally emphasized, or "modal" cues, the detail responsive children more inclined to respond to the atypical or unusual stimulus cues. Such a hypothesis is highly speculative, of course, and extensive studies would be required for its confirmation, and for the exploration of other alternatives.

Despite the very tentative nature of the hypothesis just stated, in view of the widespread interest in types of conceptual style, one is tempted to raise at least very briefly the question of possible relationships between form responsiveness and several cognitive style variables dealt with by other researchers. Is there any similarity, for example, between the form responsive-detail responsive contrast reported here, and the contrasting cognitive styles designated by

Witkin, et al. (17) as "field dependent - field independent"? Differences in the behavioral criteria for differentiating these conceptual styles would argue that this is not a strong probability, but it is an interesting question none the less. One wonders, too, whether detail responsiveness as measured in this research is related to the "analytic" conceptual style described by Kagan, Moss and Sigel (9). Here too, the probability does not seem very strong because of differences in the tasks employed, particularly in regard to the nature of the competing cues, other than details, available for use by the child. That "analytic" style and detail responsiveness probably represent different phenomena is also suggested by the fact that Kagan, Moss, and Sigel report increases in analytic responses with age, while the results of the present study indicate, for boys at least, an increase in response to form, rather than detail, with increasing age.

SUMMARY

This research was concerned with the influence of geometric form and of small details as competing stimulus cues in children's comparative similarity judgements. Attention was centered upon developmental changes and sex differences, as well as upon the scope and correlates of individual differences in cue utilization. Six comparative similarity tasks were employed, in which either over-all form or small detail could be used as the basis for judging which of four simple geometric figures was most like a standard. These tasks, along with a number of others, were administered individually to 153 children from three to eight years of age.

Preferred cue utilization was quite stable across tasks, leading to the development of a reliable (.80) over-all measure of the tendency to use form rather than detail as similarity cues. Results indicated that at all ages and for both sexes, stimulus shape was highly predominant over small details as a determinant of similarity judgements. For girls, utilization of form as a similarity cue showed essentially no change over the age range studied. Boys were less "form responsive" than girls in the early years, but demonstrated a significant increase in form utilization with increasing age, reaching a peak at the first grade, where they were significantly more form responsive than girls, then returning to the girls' level at the second grade. This final curvilinearity in the boys data represented a significant trend in the age curve, a post-hoc explanation of which is proposed in the paper.

An analysis of several correlates of individual differences in form responsiveness indicated no relationship to general ability or to perceptual acuity, but some relationship to preferential utilization of conventional meaning rather than "perceptual" qualities as criteria for judging pictorial similarities. Suggestive evidence was also found indicating a relationship between form or detail responsiveness and differential utilization of letter cues in a nonsense syllable discrimination task.

14. Vargellier, E. Détails caractéristiques et reconnaissance de formes. *Archiv. Exp. Psychol. Franciscan*, 1953, 7, 147-155.

15. Warner, E. *Comparative Psychology of Mental Development*. New York: International Universities Press, 1957.

16. Wier, B. L. *Language, Thought, and Reality*. New York: Wiley, 1956.

17. Wier, B. L., Dyk, R. B., Peterson, R. F., Coudenberg, E. A. and Mary, S. A. *Psychological Differentiation*. New York: Wiley, 1957.

REFERENCES

1. Acton, F. S. Analysis of Straight-Line Data. New York: Wiley, 1959.
2. Ammons, R. B. and Helen S. Ammons. Full Range Picture Vocabulary Test. Missoula, Montana: Psychol. Test Specialists, 1948.
3. Bruner, J. On perceptual readiness. Psychol. Rev., 1957, 64, 123-152.
4. Carroll, J. B. and J. B. Casagrande. The function of language classification in behavior. In Readings in Social Psychology. Edited by E. E. Maccoby, T. M. Newcomb, and E. L. Hartley. New York: Henry Holt & Co., 1958.
5. Gardner, R., P. S. Holzman, G. S. Klein, H. Linton, and D. P. Spence. Cognitive control; a study of individual consistencies in cognitive behavior. Psychol. Issues, 1959, 1, No. 4, Monograph 4.
6. Gibson, E. J. Perceptual development. In Yearbook of the National Society for the Study of Education. Edited by H. W. Stevenson. Chicago: The University of Chicago Press, 1963.
7. Gibson, J. J. What is a form? Psychol. Rev., 1951, 58, 403-412.
8. Goldstein, K. and M. Scheerer. Abstract and concrete behavior: an experimental study with special tests. Psychol. Monogr., 1941, 53, No. 2, (Whole No. 239).
9. Kagan, J., H. A. Moss, and I. E. Sigel. Psychological significance of styles of conceptualization. In Basic Cognitive Processes. Edited by J. C. Wright and J. Kagan. Monogr. Soc. Res. Child Develop., 1963.
10. Messick, S. and N. Kogan. Differentiation and compartmentalization in object-sorting measures of categorizing style. Percept. Mot. Skills, 1963, 16, 47-51.
11. Ricciuti, H. N. A study of abstract thinking in pre school children. In Physical and Behavioral Growth (Report of 26th Ross Pediatric Research Conference). Columbus, Ohio: Ross Laboratories, 1958.
12. Snedecor, G. W. Statistical Methods. Ames: Iowa State College Press, 1956.
13. Vurpillot, E. and H. Brault. Étude expérimentale sur la formation des schèmes empiriques. Année Psychol., 1959, 59, 381-394.
14. Vurpillot, E. Détails caractéristiques et reconnaissance de formes familières. Rev. Psychol. Française, 1962, 7, 147-155.
15. Werner, H. Comparative Psychology of Mental Development. New York: International Universities Press, 1957.
16. Whorf, B. L. Language, Thought, and Reality. New York: Wiley, 1956.
17. Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough, D. R. and Karp, S. A. Psychological Differentiation. New York: Wiley, 1962.

REFERENCES (cont)

18. Wohlwill, J. Developmental studies of perception. Psychol. Bull., 1960, 57, 249-288.
19. Woodward, E. Perception of artificial words by beginning readers. Unpublished Master's Thesis, Cornell University, 1962.

Cornell University

This research represents the second of several studies sponsored with developmental changes in children's utilization of various kinds of stimulus cues when making comparative similarity judgments, or when engaged in sorting behavior. A previous study (9) dealt with the role of over-all shape and small details as competing similarity cues in a series of matching-with-standard tasks involving simple geometric figures. The present research focused on the question of stimulus determinants of comparative similarity judgments dealing with pictures of familiar objects rather than simple geometric forms. More specifically, the study was concerned with the role of two types of competing criteria for judging the comparative similarity of pictorial stimuli: a) representational meaning, or the conventional identification of the object or class of objects represented by the "picture-as-a-whole"; and b) a variety of highly specific small details or "perceptual" aspects of the pictures, which most adults or older children would tend to regard as trivial or unimportant in the light of the alternative mode of classification available.

The reader will recognize that the distinction between the types of criteria for object classification just described corresponds very closely to the distinction traditionally made by early students of concept formation

¹This research was supported by a grant from the U.S. Office of Education, Dept. of Health, Education, and Welfare. Appreciation is expressed to Dr. Florence S. Kieckhefer and Etta Woodward for data collection, to Dr. Florence S. Kieckhefer for construction of pictorial similarity tasks, and to Mabel Lee, Marion Prentice, and Linda Johnson for assistance in statistical analyses.

Selective cue utilization in young children's comparative
similarity judgments of pictorial stimuli¹

Henry N. Ricciuti
Cornell University

This research represents the second of several studies concerned with developmental changes in children's utilization of various kinds of stimulus cues when making comparative similarity judgments, or when engaged in sorting behavior. A previous study (9) dealt with the role of over-all shape and small details as competing similarity cues in a series of matching-with-standard tasks involving simple geometric figures. The present research focused on the question of stimulus determinants or comparative similarity judgments dealing with pictures of familiar objects rather than simple geometric forms. More specifically, the study was concerned with the role of two types of competing criteria for judging the comparative similarity of pictorial stimuli: a) representational meaning, or the conventional identification of the object or class of objects represented by the "picture-as-a-whole"; and b) a variety of highly specific small details or "perceptual" aspects of the pictures, which most adults or older children would tend to regard as trivial or unimportant in the light of the alternative mode of classification available.

The reader will recognize that the distinction between the types of criteria for object classification just described corresponds very closely to the distinction traditionally made by early students of concept formation

¹This research was supported by a grant from the U.S. Office of Education, Dept. of Health, Education, and Welfare. Appreciation is expressed to Dr. Florence B. Ricciuti and Etta Woodward for data collection, to Dr. Florence B. Ricciuti for construction of pictorial similarity tasks, and to Mabel Lum, Marion Prentice, and Linda Johnson for assistance in statistical analyses.

between "abstract" and "concrete" modes of thinking (4, 13). More recent investigators have no longer found this simple dichotomy very useful, and have striven for a more precise analysis of the varying "cognitive styles" employed by children and adults (6, 14), and a fuller understanding of the perceptual-cognitive mechanisms involved in categorizing behavior (2, 5). This research was intended as another step in the direction of investigation just indicated.

The theoretical framework which provided the background for this study and the preceding one was described in some detail in the earlier paper (9). Basically, both studies reflect a central concern with the importance of further investigations of the variety of stimulus cues to which young children respond in situations where they are attempting to decide for themselves which of several alternative similarity criteria provide the most "salient" or appropriate basis for categorization of a set of objects. Such studies are considered important because they should add to our understanding of the manner in which the following factors operate as at least partial determinants of selective cue utilization in comparative similarity judgments:

- a) the intrinsic "attractiveness" or stimulus value of the cues; b) the kinds of stimulus identifications and categorizations stressed in the child's culture or sub-culture; and c) characteristics of the individual subject influencing cognitive organizational activity.

The particular impetus for both this study and the previous one (9) came from the observation (8) that many normal children show considerable interest in detailed, more or less "inconsequential" features of pictures of familiar objects, and often employ such cues as a basis for categorizing stimuli even though such a classification is incompatible with alternative categorizations based upon the conventional meaning or identity of the objects represented in the pictures. As a first step in further exploration of the scope and possible implications of children's utilization of such

stimulus properties as similarity cues, the previous study was aimed at examining the influence of small details as compared with that of over-all form in children's comparative similarity judgments of simple geometric stimuli. Interest was centered on developmental changes and sex differences over the age span from 3 to 8, and upon the scope and correlates of individual differences in cue utilization. The results indicated that stimulus shape was highly predominant over small details as a determinant of similarity judgments, at all ages and for both sexes. For girls, there was no change in responsiveness to form as a similarity cue over the age range studied. Boys, however, were less "form responsive" than girls in the early years, but became significantly more responsive to form with increasing age, reaching a peak at the first grade level, where they were significantly more form responsive than girls. In the second grade, boys' responsiveness to form decreased to the same level as that of the girls. The tendency to use form rather than detail cues was unrelated to general ability and perceptual acuity, but some evidence was found for a relationship between form or detail responsiveness and differential utilization of letter cues in a nonsense syllable discrimination task.

As already indicated, the present study was an attempt to investigate further the extent to which children respond to a variety of detailed, fragmentary, or "non-representational" attributes of pictorial stimuli as criteria for judging similarity between pictures, when such cues are in competition with common identity, or class membership as alternative criteria for classification. The general purpose was to explore the stimulus value of a rather broad range of such cues and of the competing "representational" cues, in a variety of pictorial tasks involving a matching-with-standard procedure, over the age range from three to eight years. As in the previous study, the research was concerned with developmental changes and sex differences in

utilization of the similarity criteria under study, and also with the scope and possible significance of individual differences in such cue utilization.

METHOD

Subjects

The sample used in this study was the same as that employed in the earlier research (9). The subjects consisted of 153 children, 78 girls and 75 boys, ranging from 3 to 8 years of age, representing an essentially middle class group from a public elementary school and several nursery schools in the city of Ithaca, N.Y. The total sample consisted of five age-grade groups: three-year-old and four-year-old nursery school children, along with kindergarteners, first, and second graders. A reduced sample of 125 children consisting of 13 girls and 12 boys at each age-grade level was used for purposes of statistical analyses requiring proportional numbers of boys and girls at each age level. (Further details concerning the sample may be found in the previous paper.)

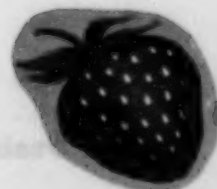
Procedure

a. Pictorial similarity tasks. A series of eight pictorial similarity tasks were employed in the study, each consisting of a standard picture and four choice pictures which were placed in a row several inches above the standard on the surface of the examining table. The pictures were printed photographically, all but one task in color, on small individual test cards. Figure 1. presents a reproduction of two of the tasks in the series, one consisting of black-and-white drawings (II-6), the other of colored prints (II-8). The original stimuli, both colored and achromatic, were selected from illustrations in children's books and educational games. It should be mentioned that the actual test pictures were about 1/3 as large again as the samples reproduced in Figure 1, and it was quite clear that children could recognize them as representing familiar objects without difficulty.

Figure 1. Samples of Pictorial Items



11-6



11-8

Figure 1: Samples of Pictorial Items

When the pictures had been arranged on the table in front of the subject, he was asked to look at the standard, then at the four pictures just above it, and to select one of the four choice pictures which he regarded as most like the standard. After he had made his first similarity choice, that particular card was removed from the table, and he was asked to judge which of the remaining three figures was most like the standard. After each judgment, the child was asked to explain very briefly how the choice figure and the standard had been perceived as alike.

The rationale which guided the construction of the pictorial similarity tasks was intended to produce sets of stimuli such that on each task, there would be at least one choice figure, and usually two, which resembled the standard in terms of common identity or class membership, but differed from it in a number of specific characteristics. At the same time, there was to be at least one, and usually two stimuli which resembled the standard picture in some structural characteristic, small detail, or color, but differed from the standard in terms of the kind of "thing" it represented.

To illustrate with the sample tasks reproduced in Figure 1, in Item II-6 the child may judge the GOAT or the DUCK as most like the standard PIG on the grounds that in either case, they are both animals. On the other hand, he could ignore either of the above choices based on common class membership, and utilize as his criteria of similarity the large black spots on the PIG and in the beard or fur collar of the KING, or the small black dots on the PIG and on the APPLE. In either case, a choice based on these similarity cues would ignore the fact that the stimuli judged as alike are grossly different in that they represent different kinds of objects.

Consider next Item II-8, in which the standard is a bright red MITTEN with yellow-white spots. The most obvious similarity choice based on common identity would be the light blue MITTEN, with the purple SWEATER offering another choice on the basis of common class membership. On the other hand,

Table 1. Description of pictures employed as stimuli in eight Pictorial Similarity Tasks.

I-2	PIG (black & white)	APPLE (Red, green spot)	BANANA (Spotted)	DOG* (frisky brown terrier puppy)
	DOG (sitting, white dalmatian, brown spots, red collar)			
I-3	TURKEY (white, tiny black spots)	GRAPES* (blue, greenish yellow stem)	CARROT (orange, green stem)	ROBIN (brownish-purple back, orange breast)
	PEAR (yellow-orange, tiny black spots)			
I-6	CHICKEN* (brown)	DOG (white, few brown spots)	BABY (crawling, yellow hair, blue shirt)	DUCK (outline drawing)
	ROOSTER (brown, yellow, blue)			
I-8	SLED* (tan, reddish-black runners)	KITTEN (brown & white, red bow)	MANTEL CLOCK (brown)	DRUM (brown, bluish-gray rims)
	WAGON (red, bluish-black tires)			
II-2	PIGEON (blue)	DUCK (standing, facing right)	TEDDY BEAR (sitting, pink ears)	BOY* (red hair, green sweater)
	GIRL (red hair, blue dress, pink skin)			
II-3	PIE (blueberry, on blue background)	APPLE* (red, dark stem)	BEE (yellow-black, wings blue-streaked)	MAN'S HAT (blue, green band)
	PLUM (blue, green leaves)			
II-6	GOAT* (outline drawing)	APPLE (black stem, tiny black spots)	KING'S HEAD (black beard, black spots on collar)	DUCK (outline drawing)
	PIG (large black spots)			
II-8	OWL (brown & white, spots, green eyes)	MITTEN* (plain blue)	SWEATER (purple)	STRAWBERRY (red, white spots, green background)
	MITTEN (red, yellow-green spots, green background)			

*Choices "keyed" in obtaining PSR-8 scores.

the bright red STRAWBERRY with white spots, has a close resemblance to the standard in terms of over-all color and visual "surface texture". The OWL also has a number of spotted surfaces like the standard MITTEN, as well as a green coloration around the eyes resembling the greenish background color surrounding the standard red MITTEN.

The remaining tasks in the series were constructed in much the same way, in the hope of sampling a rather broad variety of specific, "non-representational" stimulus properties whose role as criteria for judging similarity could thus be examined when they were "pitted against" the competing similarity criteria of common identification or class membership.¹ Table 1 presents a brief description of the main features of the stimuli employed in all eight tasks in the series.

b. Administration of instruments. The eight pictorial similarity tasks, along with a number of other items, were administered by two experienced female examiners in two testing sessions several days apart. (The other tasks consisted primarily of the geometric similarity items discussed in the previous paper.) Four of the pictorial tasks were given on the first testing day, as the 2nd, 3rd, 6th, and 8th item in the series, with the remaining four appearing in corresponding serial positions on the second day's testing. To provide a short measure of general ability level for each child, the Ammons Picture Vocabulary Test (1) was administered, with minor modifications in administration and scoring, at the end of the series on the second day.

¹Because of the exploratory nature of this particular investigation, no attempt was made to design the stimuli in such a way that the relative "strength" of different "non-representational" cues could be quantitatively assessed by pitting them against identical, or systematically varied competing cues.

ANALYSIS OF RESULTS

Reliability of Cue Utilization

In order to determine the degree of consistency in the similarity criteria employed on the eight pictorial tasks in the series, the number of first choices made on the basis of common identity or class membership for the four Day 1 tasks was correlated with the number of such judgments made on the four Day 2 items. A satisfactorily high degree of consistency across tasks (and several days' time) was revealed by the estimated split-half reliability of .80 for the total sample of 153 children. (Correspondingly reliability estimates for the five age-grade groups, from youngest to oldest, were: .40, .77, .79, .82, and .65.) Encouraging evidence for long term stability of preferred cue utilization was found in the test-retest correlation of .54 (uncorrected) between scores on four items administered to 14 children in the kindergarten group who were tested again over a year later as part of a small replication sample.

Developmental changes in cue utilization

Two types of analyses were conducted to examine the question of age changes in the stimulus cues employed by children as criteria for comparative similarity judgments: analysis of data based on over-all performance on all eight pictorial tasks, and separate analyses of response choices for the individual tasks.

a. Analyses based on over-all performance. As already indicated, it was possible to obtain satisfactorily reliable measures of children's tendency to make similarity choices on the basis of common identification or class membership, rather than on the basis of more specific, detail characteristics of the pictures. An over-all score (PSR-8) was therefore computed for all children, indicating the total number of tasks for which the first similarity choice was the stimulus most resembling the standard in terms of common

identity or class membership.² Mean PSR-8 scores were computed for boys and girls at every age-grade level for both the total and reduced sample. The results are shown graphically in the curves presented in Figure 2.

It will be noted that the curves for the total and the reduced samples are very similar, and indicate an appreciable age increase for both boys and girls in the number of "representational" first choices made on the eight tasks. In the two youngest groups, boys made fewer such responses than girls, but then showed a steeper rise in these choices than the girls, reaching a maximum score at the first grade level, while girls did not achieve this level until the second grade. It will be noted, also, that in the youngest groups, between 4 and 5 of the first similarity choices, on the average, were based on common identification or representational meaning, while at the 2nd grade level (and at the first grade level for boys only), between 7 and 8 choices, or nearly all, are so determined.

An analysis of variance of the age trends and sex differences just described was conducted on the reduced sample, with the results summarized in Table 2. It is apparent that the differences in PSR-8 scores among the five age groups were highly significant ($F = 14.90$, $P < .01$) and that the age variation for boys was not significantly different from that for girls. When a trend analysis was carried out (11) to evaluate the form of the curves, it revealed a significant linear age trend, or a significant linear relationship between PSR-8 scores and age, for both boys and girls as well as for the combined sample.

If we consider next the apparent curvilinearity in the boys data, the significant ($P < .05$) cubic component suggests that the slight "S" shaped quality of the curve represents a non-random curvilinear deviation from the

²The choices "keyed" for this PSR-8 score are indicated in Table 1. For each item, the keyed choice was the alternative most commonly selected at all ages, and chosen more frequently with increasing age.

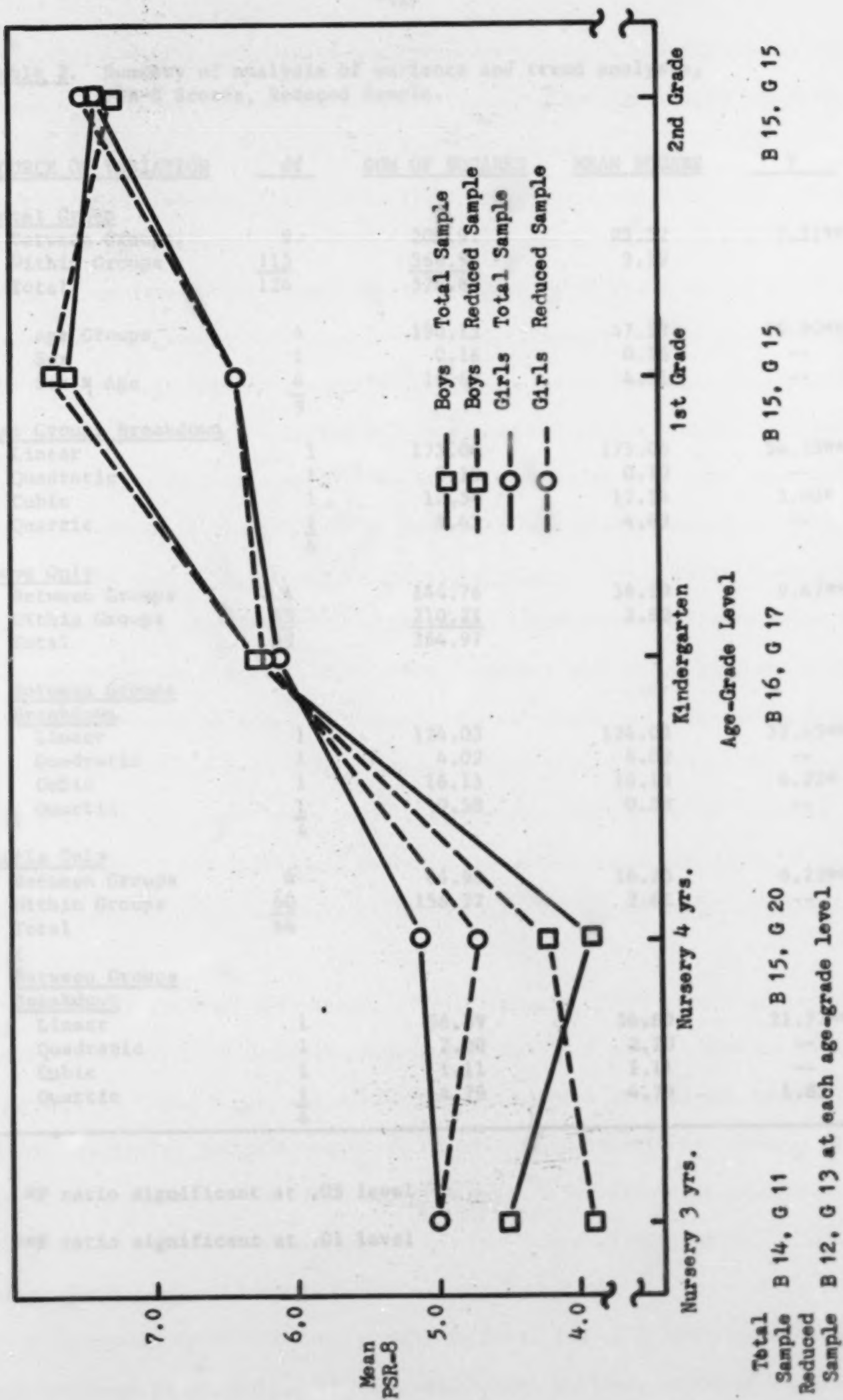


Figure 2: Mean PSR-8 Scores for Boys and Girls at Successive Age-Grade Levels
Total and Reduced Samples

Table 2. Summary of analysis of variance and trend analysis,
PSR-8 Scores, Reduced Sample.

<u>SOURCE OF VARIATION</u>	<u>df</u>	<u>SUM OF SQUARES</u>	<u>MEAN SQUARE</u>	<u>F</u>
<u>Total Group</u>				
Between Groups	9	209.91	23.32	7.31**
Within Groups	<u>115</u>	<u>366.94</u>	3.19	
Total	124	576.85		
Age Groups	4	190.13	47.53	14.90**
Sex	1	0.16	0.16	--
Sex x Age	<u>4</u>	<u>19.62</u>	<u>4.91</u>	<u>--</u>
	9			
<u>Age Groups Breakdown</u>				
Linear	1	173.06	173.06	54.23**
Quadratic	1	0.10	0.10	--
Cubic	1	12.54	12.54	3.93*
Quartic	<u>1</u>	<u>4.43</u>	<u>4.43</u>	<u>--</u>
	4			
<u>Boys Only</u>				
Between Groups	4	144.76	36.19	9.47**
Within Groups	<u>55</u>	<u>210.21</u>	3.82	
Total	59	354.97		
<u>Between Groups Breakdown</u>				
Linear	1	124.03	124.03	32.45**
Quadratic	1	4.02	4.02	--
Cubic	1	16.13	16.13	4.22*
Quartic	<u>1</u>	<u>0.58</u>	<u>0.58</u>	<u>--</u>
	4			
<u>Girls Only</u>				
Between Groups	4	64.99	16.25	6.22**
Within Groups	<u>60</u>	<u>156.77</u>	2.61	--
Total	64			
<u>Between Groups Breakdown</u>				
Linear	1	56.89	56.89	21.77**
Quadratic	1	2.20	2.20	--
Cubic	1	1.11	1.11	--
Quartic	<u>1</u>	<u>4.79</u>	<u>4.79</u>	<u>1.83</u>
	4			

*F ratio significant at .05 level

**F ratio significant at .01 level

significant linear trend which basically characterizes the data. The girls' curve shows no significant curvilinearity, although the significant cubic component is also found when the data for boys and girls are combined. It should be mentioned, finally, that at the first grade level the boys' mean PSR-8 score was significantly higher ($P < .05$) than that of the girls, while in the youngest nursery school group, the sex difference in favor of the girls approached very closely this same level of significance. There are some interesting parallels between the differential age curves for boys and girls found in the present study and in the previous research concerned with age changes in utilization of form and small detail as similarity cues (9). Similarities and differences in the two sets of data will be discussed later in the paper.

b. Analyses of responses to individual items. The results described thus far, based upon over-all performance on all eight pictorial tasks, have shown a substantial increase with age in the frequency with which pictures resembling the standard in terms of common class membership or representational meaning were selected as first choices by both boys and girls. The analyses which follow are intended to point out some characteristics of response choices made on the individual tasks in the series, with particular reference to the range of stimulus characteristics apparently serving as similarity criteria.

Some of the most interesting findings in this regard are revealed by a joint examination of the data presented in Tables 3 and 4. Table 3 shows, for each of the eight tasks, the percentage of children at each age level whose first similarity judgment consisted of the "representational" choice, (as keyed in arriving at the PSR-8 score). Table 4 indicates the proportion of subjects who selected either one of the two "non-representational" choices as their first judgment, again separately for each task.

Examining Table 3 first, it will be noted that all tasks showed a general age increase in percentage of representational choices, although there was

Table 3. Percentage of children at each age making 1st similarity choice on basis of common identification or class membership.

<u>TASK</u>	<u>STANDARD</u> <u>FIGURE</u>	<u>"KEYED"</u> <u>CHOICE FIGURE</u>	<u>NURS.3</u> <u>N=(25)</u>	<u>NURS.4</u> <u>(35)</u>	<u>KIND.</u> <u>(33)</u>	<u>1st GR.</u> <u>(30)</u>	<u>2nd GR.</u> <u>(30)</u>
<u>Day I</u>							
2	DOG	DOG	68	71	91	100	93
3	PEAR	GRAPES	36	37	67	74	93
6	ROOSTER	CHICKEN	68	83	100	94	90
8	WAGON	SLED	71	63	94	94	100
<u>Day II</u>							
2	GIRL	BOY	72	71	73	93	93
3	PLUM	APPLE	46	54	61	83	97
6	PIG	GOAT	40	31	49	66	87
8	MITTEN	MITTEN	76	46	85	93	87

Table 4. Percentage of children making either one of two "non-representational" choices as first judgment.

<u>TASK</u>	<u>STANDARD</u> <u>FIGURE</u>	<u>"NON-REPRE-</u> <u>SENTATIONAL</u> <u>CHOICES</u>	<u>NURS.3</u>	<u>NURS.4</u>	<u>KIND.</u>	<u>1st GR.</u>	<u>2nd GR.</u>
<u>Day I</u>							
2	DOG	BANANA) APPLE)	32	17	6	0	3
3	PEAR	TURKEY) ROBIN)	40	29	27	10	3
6	ROOSTER	DOG) BABY)	16	9	0	3	3
8	WAGON	KITTEN) CLOCK)	21	29	3	6	0
<u>Day II</u>							
2	GIRL	PIGEON) DUCK)	20	23	21	7	7
3	PLUM	BEE) HAT)	50	37	27	10	0
6	PIG	APPLE) KING'S HEAD)	32	49	24	10	7
8	MITTEN	STRAWBERRY) OWL)	24	54	12	7	2

considerable variation in the steepness of the age curves for the different tasks. For five of the tasks, even in the youngest nursery school group over two-thirds of the children judged similarity on the basis of common identification or class membership, suggesting that for these items representational meaning was a strong similarity criterion, and/or, the competing detail cues were not so salient as to outweigh common identification or meaning.

The remaining three items (I-3, II-3, and II-6) had considerably lower percentages of representational choices in the youngest groups, and showed the steepest age increases from 3 to 8. An inspection of the data in Table 4 suggests that for these three tasks particularly, the small structural details and color characteristics were relatively potent cues and served as the predominant similarity criteria for an appreciable number of children in the two nursery school groups as well as in the kindergarten. In the case of the "PEAR-GRAPES" task (I-3) for example, 40 percent of the youngest children ignored the GRAPES and CARROT stimuli as possible representational choices, and selected instead as most like the PEAR either the TURKEY (which had some tiny black spots, as did the pear), or the ROBIN (which had an orange breast very similar to the coloration of the pear, and a curved breast contour very similar to the contour of the pear). Similar first choices were made by 29 percent of the four-year-olds, 27 percent of the kindergarteners, and 10 percent of the 1st graders.

Proportions of children approximately equivalent to those just mentioned made the non-representational choices on item II-3, where the dark blue color of the MAN'S HAT induced many children to select it as like the PLUM, and on item II-6, in which the dark black spots were utilized by a good many children to match either the KING'S HEAD or APPLE with the standard PIG. An appreciable number of similarity choices based on non-representational cues of the sort just described were made on all tasks, particularly in the younger groups. An examination of Table 4 in conjunction with Table 1, which

describes the individual stimulus pictures, will indicate some of the particular cues involved in these judgments. Although it is sometimes very difficult to determine precisely what property of the stimulus the child is responding to, there was much evidence in the children's spontaneous verbalizations and post-hoc explanations to indicate that the cues built into the "non-representational" choices were in fact playing an important role as similarity cues.

It should again be emphasized at this point that the majority of similarity judgments, even in the younger groups, were made on the basis of common identification, or class membership. This finding is particularly evident when one considers the frequency of occurrence of various possible 1st and 2nd response-combinations on each task. With very few exceptions, in every age group the most common dual response-combination was one in which the first choice was the "representational" choice keyed in the PSR-8 score, and the second was a stimulus which could also be judged like the standard in terms of common class membership (e.g. GRAPES and BREAD given as 1st and 2nd choices to match the PEAR in Item I-3). Despite this emphasis on representational meaning or common class membership as a similarity criterion, however, the fact that an appreciable number of judgments are based on a variety of detail or color characteristics, particularly in the three youngest groups, is a finding of considerable interest. As already indicated, it was not possible, with the tasks employed in the present study, to make any systematic comparison of the "attractiveness" or stimulus value of different types of structural, detail, or color cues, since this would require placing these different cues in competition with equivalent or identical "representational" choices. Nevertheless, the findings suggest that a wide range of such "fragmentary" or partial stimulus cues may be employed by young children as dominant similarity cues. They also suggest that some children are particularly prone to make judgments on the basis of such criteria, while others respond primarily

to common identity or class membership. We turn next to the question of the possible psychological significance of these individual differences among children.

Implications of Individual Differences

The PSR-8 score which was employed in the analysis of variance previously discussed, indicated the number of tasks for which the first similarity judgment involved the commonly selected "representational" choice. In order to explore some possible correlates of individual differences in cue utilization, it was decided to use an index which reflected both the first and the second judgment made by children on four of the tasks, selected partly on the basis of steepness of the age curves for "representational" choices, partly to provide varied test content from both Day 1 and Day 2 (I-3, I-8, II-6, and II-8). The index employed (PST score) was obtained by giving maximum weight (5) to dual responses where both choices were based on common identity or class membership, and a minimum weight (0) when both choices were "non-representational", with intermediate weights being assigned according to whether single representational choices were selected as first or second responses.³

It has already been shown that there was a significant linear increase with age in the frequency of similarity judgments made on the basis of common identity or class membership, rather than on the basis of fragmentary, partial stimulus cues. The magnitude of this relationship, as expressed by Pearsonian correlations between the PST scores and age, was found to be .54 for boys, .43 for girls, and .48 for the total sample of 153 children. This is not a surprising finding, certainly, in the light of the results of many studies of concept formation and abstract thinking.

³The PST scores and the PSR-8 scores have considerable overlap as indicated by correlations of .90 and .70 found in the 1st grade and four-year-old nursery school groups.

The next question which arose in attempting to understand the nature of the individual differences in cue utilization was whether the tendency to respond to common identity, or representational meaning, which increases with age, was a function of above average intellectual capacity. It was possible to examine this question, at least in a preliminary way, by determining the correlation between PST scores and the Ammons Picture Vocabulary measures (which are reported (1) as having reasonably high validity against traditional individual intelligence test scores). Since the general ability measures employed were raw scores on the Ammons, which are highly correlated with age ($r = .83$ for the total sample), the relationship between preferred cue utilization and general ability was analyzed by determining partial correlations between PST scores and Ammons raw scores, holding age constant. The partial correlations thus obtained were .15 for boys, .19 for girls, and .19 for the total group, (with the latter value reaching the .05 level of significance). These results suggest that the tendency to be particularly responsive to representational meaning as a similarity cue was not a function of advanced intellectual development relative to one's chronological age. This hypothesis should be checked however, employing more adequate measures of intellectual status.

It was also possible to examine the relationship between PST scores and the ability to make rather fine perceptual discriminations between simple geometric figures for a different sample of 20 boys and 20 girls tested just after completing the first grade. Correlations between the PST measures and perceptual acuity scores (based upon a test described in the previous study, 9) were low and insignificant.

Finally, an analysis was made of the relationship between the PST scores and the tendency to utilize form rather than detail as preferred similarity cues in judgments of geometric figures. As already indicated in the previous report (9), small but significant correlations were obtained,

even when the effect of age was partialled out. Pearson r 's were .23 for boys, .29 for girls, and .26 ($P < .01$) for the total group, thus suggesting some degree of consistency in the tendency to respond to over-all shape rather than to small details in making comparative similarity judgments on geometric figures, and the inclination to make use of common identity or class membership rather than partial structural or color details in judgments of familiar pictorial stimuli.⁴

DISCUSSION

The fact that with increasing age both boys and girls showed a significant increase in the frequency with which comparative similarity judgments were based on common identity or class membership, and a decrease in the utilization of fragmentary details or color characteristics, is a finding which is quite congruent with the results of previous studies concerned with developmental changes in children's ability to employ abstract or "conceptual" rather than "perceptual" categorizations in sorting behavior (7, 10).

There are a number of interesting comparisons that can be made between the findings of the present research, and those resulting from the earlier study (9) which employed the same subjects, and was concerned with age changes in utilization of over-all form, rather than small stimulus details, as similarity cues. The first point of comparison is that for both sexes, the age increase in judgments based on common identity or class membership was considerably greater than that found for responsiveness to form rather than detail. (Correlations between PST scores and age were .54 and .43 for boys and girls respectively, while the corresponding values for form responsiveness and age were .29 and -.02.) A plausible explanation of these differences may be found in the fact that in the case of many of the pictorial

⁴Supporting these findings was a correlation of .36 ($P < .05$) found between PST scores and form responsiveness scores for the replication sample of 40 first graders.

tasks, selection of the choices given the highest weight in the PST score is facilitated if the child can make the inferences required to recognize that two stimuli belong in a common category (e.g. WAGON and SLED both "toys" or "to ride on"). On the other hand, making the "form" choice on the geometric similarity tasks requires merely that the child recognize the similarity in outlines of two simple geometric forms which are essentially identical, except for the differences in the small detail qualities. In short, it seems reasonable to assume that on many of the pictorial tasks a greater degree of inference is required, and specific stimulus differences of considerably greater magnitude must be ignored, than is the case with the geometric stimuli. Since it is well known that the capacity to make inferential judgments shows rather marked developmental growth in children over the age range studied, this fact may provide an explanation for the considerably steeper age curves found for utilization of common identity or class membership similarity criteria in judgments of the pictorial stimuli.

Despite the differences just discussed in the magnitude of the relationships found between age and preferred cue utilization on pictorial as compared with geometric stimuli, there are some rather striking parallels in the differential patterns of age changes for boys and girls revealed by the two sets of data. In both cases, boys in the youngest groups make considerably less frequent use of characteristics of the "stimulus-as-a-whole" than do the girls (i.e., they are less form responsive, and make fewer choices based on common identity or class membership). This is followed in both studies by a more rapid increase in the utilization of such stimulus characteristics for boys than is true for the girls, with the boys reaching a peak in this regard, which exceeds considerably that of the girls, at the first grade level. Finally, boys and girls have essentially equivalent means by the second grade. (It should be mentioned that there were no ability differences between boys and girls in the various age-grade groups.)

In the previous paper, it was suggested that the differential age pattern for boys and girls might be explained by the hypothesis that girls respond earlier and more markedly than boys to the emphasis which the culture places upon the over-all shape of things, and upon the identification of simple geometric forms; hence they manifest a stronger selective utilization of such cues in the early years. Considering the parallel sex differences in the youngest groups shown in the present study, one wonders if a similar explanation might be hypothesized in this instance. Young children generally learn to be increasingly concerned with the recognition and identification of objects-as-a-whole in their environment, and in the process, they learn which cues or stimulus details are necessary (or "criterial") for such identifications (3, 12), and which are irrelevant. It is conceivable that girls respond earlier and more readily than boys to this emphasis upon object identification, and hence showed a more marked selective utilization of common identity or class membership, rather than "inconsequential" details or color cues, in making comparative similarity judgments on the familiar pictorial material employed in this study.

It was suggested in the previous paper that the boys' markedly heightened response to form cues at the first grade level, which significantly exceeded that of the girls, might be attributable to the greater emphasis reportedly placed by teachers upon form discrimination and reading readiness training for first grade boys than for girls. Whether this comparatively more intensive training in form recognition might have contributed to the first grade boys' sharply increased responsiveness to the "picture-as-a-whole" in terms of common identity or class membership, which at this level was significantly greater than that of the girls,⁵ is a speculative matter. However, it deserves some consideration nonetheless.

⁵It should be pointed out that in the small replication sample of 20 boys and 20 girls tested in the summer following completion of first grade, the mean PST scores for boys and girls were almost identical, so that the significant sex difference discussed above was not replicated.

The fact that the utilization of small details or color properties as preferred similarity cues declined with increasing age in the present study, may appear at first glance to conflict with the findings of a recent investigation by Kagan, Moss, and Sigel (6). These authors employed a series of pictorial triad items, on which children were asked to select two of the three figures which were alike or went together in some way. Children's responses were coded as "analytic" if the similarity judgment was based on common specific details, "relational" if based on some type of functional relationship between the two figures selected, or "inferential" if based on common membership in a broad category. The point of interest here is that the authors reported that "analytic" responses, based on detail similarities, tended to increase from the first to the sixth grade, while "relational" responses decreased. The apparent discrepancy between these findings and those of the present research is probably explainable, at least in part, by the fact that the alternative or competing bases for classification were quite different in the two studies. In the present investigation, detail cues were "pitted against" common identification, or common class membership, a similarity criterion very much like that represented by the "inferential" responses described by Kagan, Moss, and Sigel (6). In the latter study, however, since the authors were particularly interested in measuring individual differences in analytic responses as an indicator of children's cognitive style, their test items were constructed so as to minimize the likelihood of inferential responses occurring, since such choices would constitute a preponderance of the responses with increasing age. Hence, the choices for the most part had to be either analytic or "relational", so that the task posed for the child represented a very different one from that employed in the present study. It was under these task conditions, then, that analytic responses based on similarity in details increased with age. The different findings just discussed represent a good illustration of the difficulty

involved in attempting to arrive at generalizations concerning preferential cue utilization, since particular results are always so highly dependent on the nature and strength of the competing or alternative cues available to the subject. As a matter of fact, a profitable line of future investigation might well focus on systematic variation of particular types of stimulus cues which are pitted against systematically varied alternative modes of classification.

The possible psychological significance of individual differences in preferred cue utilization is a question which requires much further study. That there are relatively wide reliably measurable individual differences is quite clear. These differences do not appear to be explainable in terms of above average or below average intellectual capacity, or to perceptual acuity. The fact that there was some degree of generality between preferred utilization of form rather than detail as similarity cues in judging geometric stimuli, and the tendency to judge pictorial similarities primarily on the basis of common identification or representational meaning, suggests that both of these measures may reflect a more generalized tendency to be particularly responsive either to the culturally emphasized, common identifications and classifications of objects and things in the environment when engaged in comparative similarity judgments and categorizing behavior, or to the less commonly stressed, less obvious, or unusual characteristics of stimuli.

SUMMARY

The present investigation was concerned with the role played by detailed, fragmentary, or "non-representational" stimulus cues as compared with that of common identity or representational meaning, in a matching-with-standard procedure where these two types of stimulus attributes constituted competing criteria for judging the comparative similarity of familiar pictorial stimuli.

Eight pictorial similarity tasks were given to 153 children from 3 to 8 years of age, the sample also employed in a previous study of form and detail as determinants of similarity judgments of simple geometric stimuli.

Children's preferred type of cue utilization could be measured with a reasonably high degree of reliability (.80). Analyses of age trends and sex differences in preferred similarity cues, using an over-all measure based on all eight tasks, revealed substantial, significant linear increases with age for both sexes in the use of common identity or class membership as the dominant similarity criterion. This relationship with age was somewhat stronger for boys ($r=.54$) than for girls ($r=.43$), who made considerably more use of common identity or representational meaning than boys in the youngest groups. By the first grade level, boys had reached a point where they made significantly more use of such stimulus attributes than did the girls.

There was a considerably stronger relationship between age and increasing use of attributes of the "stimulus-as-a-whole" as similarity criteria in the present study than in the previous investigation, where increasing use of form cues was found only for boys. However, there were some striking parallels in the differential patterns of age changes found for boys and girls in the two studies. Although common identity or representational meaning was the main similarity criterion at all ages, a considerable proportion of the responses, particularly at the younger age levels, were made on the basis of a wide variety of detailed, non-representational stimulus characteristics. The tendency to make judgments based on one type of similarity criterion rather than the other seemed to be unrelated either to general intellectual level, relative to age, or to perceptual acuity. However, there was some degree of generality across both geometric and pictorial stimuli, in the inclination to base similarity judgments on characteristics of the stimulus-as-a-whole, rather than on small details or fragmentary cues. The possible psychological significance of individual differences in these different modes of cue utilization is a question which warrants considerable further research.

REFERENCES

1. Ammons, R. B. and Helen S. Ammons. Full Range Picture Vocabulary Test. Missoula, Montana: Psychol. Test Specialists, 1948.
2. Bruner, J. S., Goodnow, J. J., and Austin, G. A. A Study of Thinking. New York: Wiley, 1956.
3. Gibson, E. J. Perceptual development. In Yearbook of the National Society for the Study of Education. Edited by H. W. Stevenson. Chicago: The University of Chicago Press, 1963.
4. Goldstein, K. and M. Scheerer. Abstract and concrete behavior: an experimental study with special tests. Psychol. Monogr., 1941, 53, No. 2 (Whole No. 239).
5. Inhelder, B. and Piaget, J. Le genèse des structures logiques élémentaires: classifications et sériations. Neuchâtel: Delachaux & Niestle, 1959.
6. Kagan, J., H. A. Moss, and I. E. Sigel. Psychological significance of styles of conceptualization. In Basic Cognitive Processes, Edited by J. C. Wright and J. Kagan. Monogr. Soc. Res. Child Develop., 1963.
7. Reichard, S., Schneider, M., and Rapoport, D. The development of concept formation in children. Amer. J. Orthopsychiat., 1944, 14, 156-162.
8. Ricciuti, H. N. A study of abstract thinking in pre school children. In Physical and Behavioral Growth (Report of 26th Ross Pediatric Research Conference). Columbus, Ohio: Ross Laboratories, 1958.
9. Ricciuti, H. N. Geometric form and detail as determinants of comparative similarity judgments in young children. Cornell University, Mimeo. 1963.
10. Sigel, I. E. Developmental trends in the abstraction ability of children. Child Develop., 1953, 24, 132-134.
11. Snedecor, G. W. Statistical Methods. Ames: Iowa State College Press, 1956.
12. Vurpillot, E. Détails caractéristiques et reconnaissance de formes familières. Rev. Psychol. Française, 1962, 7, 147-155.
13. Werner, H. Comparative Psychology of Mental Development. New York: International Universities Press, 1957.
14. Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough, D. R. and Karp, S. A. Psychological Differentiation. New York: Wiley, 1962.

THE ELEMENTS OF GRAPHONOMY

Charles F. Hockett
Cornell University

(This is a private document prepared for the use of the Reading Research Group at Cornell University. Appearance in this form does not constitute publication, and materials are not to be cited from this document--except by members of the Reading Research Group--without special permission.)

Graphonomy is the branch of anthropology devoted to the study of writing systems and their history, as linguistics is the branch devoted to the study of language.

The words language and a language are used herein in their technical senses, as more or less equivalent to what the man on the street calls "spoken language." The word writage will be used for the whole human institution of writing systems; and a writage is used in place of the clumsier expression "a writing system."

I. Similarities and Differences Between Writage and Language

Leonard Bloomfield (in Language, 1933) begins his empirical study of language with the following parable:

Suppose that Jack and Jill are walking down a lane. Jill is hungry. She sees an apple in a tree. She makes a noise with her larynx, tongue, and lips. Jack vaults the fence, climbs the tree, takes the apple, brings it to Jill, and places it in her hand. Jill eats the apple.

...

If Jill had been alone, she might have been just as hungry and thirsty and might have seen the same apple. If she had sufficient strength and skill to get over the fence and climb the tree, she could get hold of the apple and eat it; if not, she would have to stay hungry. The lone Jill is in much the same position as the speechless animal. ...

Of course, it is important for Jill's welfare that she get the apple. In most instances it is not a matter

of life and death, though sometimes it is; in the long run, however, the Jill (or the animal) that gets the food has far better chances of surviving and populating the earth. Therefore, any arrangement which adds to Jill's chances of getting the apple, is enormously valuable for her. The speaking Jill in our story availed herself of just such an arrangement. She had, to begin with, the same chance of getting the apple as had the lone Jill or the speechless animal. In addition to this, however, the speaking Jill had a further chance which the others did not share. Instead of struggling with the fence and the tree, she made a few small movements in her throat and mouth, which produced a little noise. At once, Jack began to make the reactions for her; he performed actions that were beyond Jill's strength, and in the end Jill got the apple. Language enables one person to make a reaction when another person has the stimulus.

Our purposes here are not Bloomfield's; he wished to compare the possession of language with its lack, while we wish to compare the workings of language with those of writage. To this end, we present an additional parable:

Jack and Jill live in a house behind which there is an apple tree. Jill notices one afternoon, while Jack is away at work, that the apples are ripe. She decides that they should be picked so that she can make some applesauce. The task is beyond her physical powers; or else she has to go to her bridge club, and will not be back home until dinner time. She cannot wait until Jack gets home and tell him by word of mouth that he must pick the apples. Nor does she have a telephone. But there is no particular problem, since Jack and Jill belong to a literate community. Jill uses a short wooden wand on a bit of white stuff, which she places on the kitchen table, and then she goes off to her bridge club. An hour later, Jack gets home from work, finds the white stuff, and proceeds to go out and pick the apples. Jill gets the apples not only

without expending her own energy to pick them, but even without using her lungs and lips to speak to Jack.

Language accomplishes things that cannot be done by speechless animals, in that it mediates between a practical ("non-speech") stimulus on one person and a practical response from someone else. Writage accomplishes this same kind of result; furthermore, it achieves the result under some circumstances in which language could not, or in which language would achieve the result only clumsily or awkwardly.

The differences between language and writage can be listed and discussed in more detail; the following list is doubtless incomplete, but at least these differences are of importance:

(1) The use of language involves no special apparatus, only parts of the human body and features of the environment (the air, sound waves) that are omnipresent and free. Writage involves, in the normal case, the use of artifacts (pencil and paper, pen, ink, clay tablets, etc.). True enough, there are marginal cases in which writage is brought into play without any specialized artifacts: one may write with one's toe, or a piece of driftwood, in the sand of a beach, or may trace letters with one's finger on someone else's back; but, as stated, these are marginal. We should therefore expect (and the historical evidence bears out the expectation) that the development of writage should have come about long after our ancestors had been using tools of various sorts. Indeed, the earliest archeological evidence of the making of marks on flat surfaces (upper paleolithic cave-wall art, inscribed pebbles)

precedes by perhaps ten thousand years the earliest surviving evidences of true writing.

(2) The use of language involves direct and immediate energetic action, up to the time the speech signal strikes the auditory nerve of the hearer. That is, the energy that a speaker puts into the sound waves he produces is just the energy transmitted to the hearer's ear. The use of writage is indirect and not immediate. The actions of the writer produce a relatively permanent configuration in the environment; subsequently, whenever there is light about, that configuration will reflect and refract the light to yield its own image on the retina of any observer. In language, the speaker triggers the hearer. In writage, the writer triggers the environment, and the altered environment triggers the reader.

(3) Stemming directly from the foregoing is the following: The relative permanence of the external factors involved in writing enable it to span time and space as language cannot. For speech, either a receiver is within range at the proper moment or the signal is lost forever. For writage, one may read at one's leisure. Time is spanned in our parable, or whenever someone writes a memorandum to himself and reads it an hour, a day, or a year later, or whenever someone incises wedges in a clay tablet and someone else, several thousand years later, examines the result. Time is spanned when one puts down two three- or four-digit figures to multiply them, since even in this case the use of notation relieves the computer from the burden of remembering intermediate steps of the computation. Space is spanned in that a written message, if placed

on sufficiently light material, can in time be carried far beyond earshot of the writer. Another aspect of this is that a speaker normally knows who is hearing him (barring eavesdroppers), whereas a writer quite often does not know who will read what he writes.

This external spanning of space and (particularly) of time compares with the internal spanning of time that we call memory. Both written material and memory are instances of what can also be called information storage or records.

(4) The portion of the body with which we speak (the respiratory tract and the upper portions of the alimentary canal) can be collectively called the "organs of speech." It is known that in a few cases different motions of the organs of speech can produce acoustically indistinguishable sounds. One example is so-called "r-color," which some speakers of American English produce by curling the tip of the tongue up and back, but others produce by placing the tip behind the lower teeth and bunching the middle part of the upper surface of the tongue up and slightly forward. One cannot tell by ear whether a given speaker is using one or the other of these articulations. In general, however, there is a close approach to a one-to-one relationship between articulatory motions and the resulting sounds. This means that the description of speech sounds in terms of the articulatory motions that produce them is, with only marginal uncertainties, not only useful, but probably the most useful way to describe speech sounds.

It seems proper to at least raise the question of whether the same sort of situation holds for writing. There would seem to be a

very wide variety of ways in which one and the same written symbol can be produced. We may use either hand, with any of a number of writing implements against any of a number of surfaces; or we may type, or set type from a job case; or in extreme cases we can write by bending glass tubing or even by properly piloting an airplane which is emitting a column of smoke.

The act of writing would thus seem to be a special kind of drawing, or tube-bending, or plane-piloting, or the like; but the production of speech is not a special kind of anything else. The activity of speaking, in other words, seems probably to have a much more primary and independent status from other activities than does the activity of writing.

Of course, this difference and our emphasis on it does not in any way imply that the bodily motions and feedback adjustments involved in writing (or in learning to write) should not be studied. In any given culture (e.g., our own), the educational tradition puts children through certain specified paces in learning to write, and these basic writing-motions thus have life-history precedence over alternative varieties of writing-production that may be learned later.

Theoretically, one might expect that ways of producing speech other than by articulation could be developed, just as ways of producing sequences of letters other than with hand, pen, and paper have been. In fact, during recent years two such derivative ways have been invented. One is the voder: a trained operator works certain keys that turn certain electronic producers of clicks, buzzes, and hums on and off, in such a way as to produce a faintly

recognizable facsimile of speech. The other is the pattern playback: one first draws an artificial spectrogram on a strip of transparent plastic, then feeds this artificial spectrogram into the machine, which converts it into sound. Older mock-ups of speech are less deviant from ordinary articulation: laryngectomized patients have learned to swallow air and then to use stomach air instead of lung air for articulation, some bit of scar tissue doing the normal job of the vocal cords. All of this leaves me with the basic conviction expressed in the second paragraph above.

(5) The hearing of speech differs in one important way from the hearing of other sounds. Any sound that human ears can hear (and some that they can't hear) can be described physically in terms of energy-distribution at various frequency-levels and the changes of this distribution in time. (Ears seem, in general, to ignore differences of phase relationship.) In the normal case, the ear and brain extrapolate from the spectrum to interpret a fundamental (whether there is energy at that frequency or not), interpreted as pitch. Relative to this fundamental the distribution of energy at various frequencies gives us tone quality. The overall amount of energy, influenced in a complicated way by the spectrum, gives us volume. Changes of these factors with time yields rhythmic contours. Most or all of these go on in most kinds of hearing. In hearing speech, we pay attention also to the distribution of energy at certain absolute frequency levels (not levels relative to the extrapolated fundamental), and this supplies us with vowel color. It is known that vowel color tells us what vowels or diphthongs a speaker is producing, and also does a good portion of the job of

identifying his consonant articulations. The vowel color of English a (as in father) is approximately the same for a male and for a female speaker, despite the difference in fundamental. We do not usually hear vowel color except in listening to speech--not, for example, in listening to non-vocal music, save in the jazz instance in which a brass player, by moving a hat from the bell of his instrument, produces the effect waa-waa. The sonovox is an apparatus for adding vowel color to sounds produced other than by speech: this is how the foghorn on radio used to be made to say "B O." There is serious doubt that any animal communication system makes use of vowel color differences; even dogs may well ignore vowel color in responding to human commands.

What this means is that the hearing of speech is a special kind of hearing. There seems to be no comparable special feature for the seeing of writage. We see a letter or a written word as we see an inkwell or a tree or anything else; as a figure against a ground. The one arbitrary and specialized feature is that in seeing writage there is a prescribed sequence for looking at the various parts of what is spread over an area; left to right in rows, and top to bottom; or the reverse of this; or various other ways. There is no such prescription for looking at a landscape or for a picture of a landscape.

Now this prescribed "order of looking" is, for a child, an arbitrary procedure that is not necessarily learned easily. Yet we may suspect that animals could be trained to such an order of looking, just as human children are; there seems to be much more doubt that

animals could be trained to react to differences of vowel color.

These guesses would not be hard to test experimentally. If they are confirmed, then we can conclude that writage, in this connection, simply makes special use of a biological capacity that is quite widespread, whereas language requires something that is uniquely human.

(6) Primates have hands, prehensile organs that do much of the manipulating of the environment that many land mammals do with the mouth. The use of the vocal-auditory channel by language and by its various primate parallels and precursors is consonant with this: the mouth has been relieved of some of its classical burden of manipulation; at the same time, messages can be transmitted and received without interrupting what hands and eyes are doing. Thus the vocal-auditory channel is a convenient one when primates get around to performing collective tasks; the activities of the participants can be coordinated by vocal-auditory communication without requiring the activities to be interrupted.

Writage, particularly the act of writing with the relatively clumsy devices that were the earliest to be used, requires a much more thorough involvement. Writing-implements are tools; it is the hands that wield them, and the eyes must watch what the hands are doing. A group of workers who must act together to move some heavy and delicate object could hardly coordinate their actions by passing notes back and forth.

(7) Bloomfield makes an important distinction between displaced speech and speech which he does not label, but which we shall call immanent. Roughly, displaced speech is talking about something that is remote from the setting in which the talking occurs: "I had a

headache yesterday"; "Take these pills beginning tomorrow morning"; "I think there is an apple orchard beyond that hill"; and the like-- versus "Please pass that sugar to me"; "Are you all right?" and so on.

Observations made so far reveal no communicative behavior among primates (other than Man) which could be classed as displaced. For instance, a gibbon, finding food, emits a certain call that informs the other members of the band that he has found food. The other gibbons know where the food is only by knowing where the call is coming from. The gibbon who has found the food does not rejoin the others and tell them where it is. Displacement occurs among bees; but for primates, or even mammals, it seems to be uniquely human.

Obviously not all speech is displaced, but it can be so freely under almost any conditions. Writage, on the other hand, is almost always displaced. This stems, at least in part, from consideration (6) above: in order to write, one usually has to drop other activities, so that anything one writes about is for the moment relatively "remote" from one's activities of the moment (the writing). It stems also from considerations (2) and (3) above.

(8) Also stemming from (6) is the fact that writage, as a channel for communication, is typically thinner or less fluent than language in its context. As we speak with one another, we also communicate via tone of voice variation, gestures, subgestural body motions, and so on. Either the action of writing or that of reading requires considerable focusing of attention, which tends to mask out any messages being transmitted in these other channels. Often enough, we write or read when there is no one about, or when those

in our vicinity have little or nothing to do with what we are writing or reading about.

There has been much loose talk, and perhaps some tight experimentation, centering on "eye-mindedness" versus "ear-mindedness"; the notion being that some people find it more agreeable and economic to take information in through the eye, via writage, while others prefer or are more efficient at receiving information through the ear, via language. It should be pointed out that a personal preference for reception via the ear may not indicate any such thing as "ear-mindedness"; it may indicate a preference for seeing, smelling, and on occasion even touching the person who is speaking to one-- a preference for a broad bundle of channels, with opportunity for immediate give-and-take response, rather than the narrow and depersonalized channel used by most writage.

(9) In language, the transmitter (in the first instance) sets the tempo; the receiver has to sink or swim, unless he can induce the transmitter to alter the tempo. In writage, the rate of transmission and the rate of reception are independent. The writer transmits as slowly or rapidly as he wishes; the reader, later on, reads as slowly or rapidly (or repeatedly) as he wishes.

(10) In speaking, once something is said it has been said-- the speaker cannot take it back. There are, indeed, various established "cancel" signals, by which we tell those who are listening to us that something we have actually said was "wrong" and that we want to retract and replace it by something else.

Furthermore, these "cancel" signals work so well that most of the

time we are quite unaware of the enormous amount of hemming and hawing and overt "editing" that most of us indulge in in normal conversation. But even if a hearer plays the game of accepting the cancellations, he has heard the "unintended" bits too, often out of awareness--and they have given him information of one sort or another.

In writing, which is very often much more deliberate, the writer can indulge in any amount of trial-and-error, part of it on paper, and then transmit to others only the final edited form.

This is one of the factors that underlies the development of various literary styles, and produces in some written material a smoothness and elegance (by specific cultural standards) lacking in all but the most expert formal rehearsed speech. At the same time it renders "slips of the pen" very much rarer than "slips of the tongue," which in turn largely eliminates from writage certain important sources of information present in language.

(11) Speech takes place in time; it is unidimensional and unidirectional. Writage involves an array of marks on a two-dimensional surface; despite the customary convention by which the marks are to be looked at in a specific sequence, it remains true that a two-dimensional surface is involved, and that one may devise ways of making use of the two-dimensionality.

(12) Language is characterized by duality of patterning.

Any utterance in a given language, aside from pauses and hesitations and hems and haws and other such marks of the work of generating speech, consists wholly of an arrangement of minimum elements called phonemes, of which the language supplies a fixed finite stock. These phonemes are describable in terms of

articulatory motions. The differences between phonemes are absolute: if I say either pit or bit, then the first phoneme I produce is either p or b, and what you hear will be identified the one way or the other--there is no third alternative, except an inability to tell which was said (because of noise), and this third alternative is a breakdown of the system, not a possibility within the system. Phonemes do not mean anything. Their sole function is to distinguish utterances from one another.

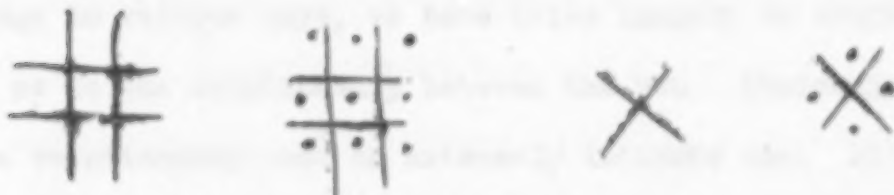
At the same time, any utterance in a given language consists wholly of an arrangement of minimum elements called morphemes, of which the language supplies a finite (though larger and not so fixed) stock. Morphemes are the minimum elements that carry meaning or that have a semantic function: i.e., such morphemes as boy or table or the un- of unproductive or the -s of boys carry meanings, while such a morpheme as and, while it has no meaning of its own, has a semantic function in that match and book means something different from match or book and from match book.

The arrangements in which phonemes occur in a particular language constitute the phonology of the language. The arrangements in which morphemes occur relative to each other constitute the grammar of the language. Obviously any morpheme, in any one occurrence, is represented by some arrangement of phonemes: the pattern of representation is a sort of code, called the morphophonemics of the language.

Although I have not yet seen any reliable account of an animal communication system incorporating this principle of duality, there

are some human systems more recent than language that do. Telegraphy with Morse code does, for example: the analog of phonemes is the dots, dashes, and spaces of various durations, while the analog of morphemes is the arrangements of dots and dashes to which meanings are assigned--thus two dots means the letter "I".

A few writages have been marked by duality; for example, the Ogam script once used for writing Old Irish. Similarly, there is a familiar schoolboy code-writing that involves duality. One obtains twenty-six graphic shapes from the figures



yielding J, U, L, J, U, >, >, A, etc. One assigns each of these to represent a letter of the English alphabet; thus each is a morpheme-analog. But the shapes themselves are easily described as composed of two, three, or four line segments, perpendicular to each other, vertical or tilted at forty-five degrees from the vertical, and with or without an included dot. These constituent strokes, which serve wholly to keep the morpheme-analogs apart, have no meaning of their own, and are the phoneme-analogs.

Most writages do not incorporate duality of patterning in more than a marginal way. In some type-fonts of the English alphabet, the four lower-case letters p, q, b, and d could be taken as a subset built out of two or three constituent marks in varying arrangements, but as soon as we try to push such an analysis to the point

of providing for all 26 letters (times two for lower-case vs. capital, and so on), we soon find that the analysis is terribly complex. That is, we find that we are forced to recognize almost as many graphic constituents as there are whole letters to be analyzed--or even more. Significant duality is a pattern by which a relatively small number of phoneme-analogs serve to differentiate a relatively much larger number of morpheme-analogs.

II. Functional Relations Between Writage and Language

In the foregoing, though we have been comparing writage and language in various ways, we have tried largely to avoid any implications as to the relationship between the two. Obviously there is such a relationship--and an extremely intimate one. All the evidence we need is to note that a literate person can read aloud (that is, transduce written material into linguistic form) or take dictation (that is, perform the inverse transduction). It is little wonder that in most literate communities the distinction between writage and language is not clearly recognized--as when, in our own community, most people speak of "spoken language" and "written language," implying that there is but a single system but with two different channels. There are other historic reasons for the lack of clarity. When the human race was learning to speak (and when any child is learning to speak) it did not have the communicative means to talk about the communicative system that was developing (as the child does not). Writage is much more recent in human history than language, and every child who learns to write and read at all acquires these skills after he has acquired the language of his

community. (To this there is an occasional pathological exception.)

This means that as writing and reading habits are acquired, by the species or by the individual, the operations involved can be talked about. Talking about writage is an older tradition than talking about language; given the very intimate relationship of the two, the development of clearly differentiated terminology has been a hard task.

The error implied by our general habit of referring to "spoken language" and "written language" is not the error of assuming a more intimate relationship between writage and language than exists, but rather that of oversimplifying that relationship.

Let us consider, for a moment, what would have to be true in order for the terms "spoken language" and "written language" to be accurate.

Suppose that we devised a set of written symbols exactly equal in number to the number of phonemes in a language, and assigned them in a one-to-one fashion to the latter. Then, since every utterance in the language consists of an arrangement of phonemes drawn from this stock, we should be able to represent every utterance in writing, without ambiguity, and should be able to say aloud anything that the rules of the writage allowed us to write. We would have a sort of isomorphism between the minimal distinctive units of speech in the language, and the arrangements in which they occur relative to each other, and the minimal symbols of the writage, with the same rules of arrangement. Except for the difference between the sounds of the phonemes and the appearances of the graphic symbols,

which would require separate description, any statement that held for the language would hold for the writage, and vice versa. A single description could apply to both.

Certain differences, of course, would remain: namely, those that turn directly on the difference between the two channels. This means virtually all of the differences itemized earlier, save possibly the last (a writing system worked out to be isomorphic with the phonological system of a language could be claimed to have the same duality involved in the language itself). It could still be argued that the two channels involved a single underlying system, at least in the sense that a mathematician can claim that the positive integers written "1, 2, 3, ..." and the positive integers written "I, II, III, ..." are essentially the "same" system (either notation satisfied the postulates for positive integers to "within an isomorphism," and this is all that concerns the mathematician).

Theoretically, this same isomorphism between a writage and a language could be achieved if some group of people taught themselves a set of phonemic habits that would be isomorphic with ordinary English spelling.

It is a fact verging on the empirical that the first way of attaining isomorphism between a writage and a language has been followed more than once, and that it is not too difficult to do, whereas the second way seems almost beyond human capacity. An experimental set-up to test the second part of this statement would be cruel beyond endurance, or else would fail: pronunciation habits go so deep that adaptation to the invented system would be

incomplete--intonation, not represented orthographically, would be added, and various other features of pronunciation not planned for the artificial language would creep in and assume distinctive functional (phonemic) roles.

This suggests something that the linguist wants to insist on: that even in a set-up where a language and a writage were virtually isomorphic, the language would in some senses be more basic than the writage. That is, it would remain true that children would learn the language first, and only later learn the writage; it would be true that sequences of symbols in the writage would not carry their meanings directly, but indirectly, mediated by the language itself. Sometimes we linguists want to go further, and say that the literate individual, reading material in the writage silently, is converting the arrays of graphs into speech sounds "inside his head" and "understanding" only on the basis of the speech sounds. But this is purely introspective and can have, at most, rhetorical value in underscoring the socio-cultural priority of language over writage.

Most actually used writing systems do not have the property of isomorphism with a language of the kind described above. Although a literate person can in general take dictation, if he does so he necessarily omits certain communicatively relevant phenomena: the transduction from speech to writing is a reduction. Contrariwise, although a literate person can read aloud, the speech he produces under the guidance of what he sees fails to reflect certain visual features that are beyond all doubt communicatively

important. Even if we constrain our consideration to that subset of features of a language that are represented in the writage, and vice versa, the correspondence between the two is not one-to-one, but (for most writing systems) replete with irregularities and ambiguities.

In the light of this, it turns out that while a language can be quite satisfactorily described without reference to any particular writage (though the linguist will invent a special one, a phonemic transcription, as a matter of convenience), it is quite impossible to describe the detailed workings of a particular writage without using the associated language as a frame of reference. This fact, again, underscores either the more basic nature of language or the living prejudice of linguists in favor of the more basic nature of language.

III. Types of Writages; Typical Complications

If we set aside any features of a writage which correlate with nothing in the associated language, then writages can be classed according to the ways in which they correlated with their associated languages, and the irregularities and complexities of specific writages can be similarly categorized. This is done in sections 62.2-3 of my Course in Modern Linguistics, which is being separately hectographed to accompany the present statement.

IV. The History of Writage

A brief outline history of writage is covered in section 62.4 of the source cited above; this, also, is being separately hectographed.

Brief comments on "graphemics" and the sociology of writing will follow.

ANALYSIS OF ENGLISH SPELLING

C. F. Hockett

Cornell University

Part I. ANALYSIS OF GRAPHIC MONOSYLLABLES

For our purposes, a graphic word is a combination of letters separated from any others by spaces or hyphens (except for the hyphen at the end of a line, here ignored).

Some graphic words contain marks other than letters--apostrophe, diacritics. Such words are ignored here. Some graphic words are distinguished from others by the distinction between capital and lower case (e.g., the shape "house," representing the ordinary word, and the shape "House," representing a personal name). This distinction is ignored here.

There are twenty-six letters, A through Z. Five of them are vowel letters: A E I O U.

A graphic monosyllable is a graphic word that includes not more than two vowel letters, and in which no consonant letter intervenes between successive vowel letters.

In the following, English graphic monosyllables are classed (1) according to the arrangements of letters in them, and (2) in terms of their pronunciation. The aim of the classification is to discover those rules by which pronunciation can be predicted from spelling. A rule is an assertion that if the spelling is such-and-such, the pronunciation is so-and-so. For example: "If the first letter of a graphic monosyllable is B, and if that letter is followed by a vowel letter, the pronunciation of the word will begin with the English consonant phoneme /b/." A graphic monosyllable whose

pronunciation is not predictable from its spelling by such rules contains or constitutes an irregularity. Rules can be formulated in several ways. By allowing rules to become very complicated, we can minimize the number of irregularities. By allowing a larger number of irregularities, rules can be rendered simpler. Between these extremes there is some optimum or near-optimum compromise. A complicated rule that covers only a single graphic monosyllable is obviously not helpful. A simple rule that covers all but a very few of a large number of common (frequent) graphic monosyllables is obviously useful.

In this presentation we do not actually select and present a specific set of rules. Instead, we present all the relevant data on which the selection should be based, except (1) actual frequency figures for the graphic monosyllables covered, and (2) a very small number (about 100) of very rare graphic monosyllables.

In theory, we should be able to distinguish at least roughly between two kinds of written words: (1) habit-forming words, and (2) analogy-following words. The former words are those that the English-speaking child almost inevitably knows (in spoken form) before he first encounters them in learning to read. They are the words in terms of which he builds up the associations between spelling and sound that form the basis of literacy. The latter words are those that the child does not encounter in reading until later, and that he may very well never have learned in oral form. His pronunciation of such a written word, when first encountered, will be based on the analogical extension of his spelling-to-sound habits as established by the first set. Thus if the child (or an adult) sees an unfamiliar written word beginning with B, he will read

it off with an initial /b/ because habit-forming words have established that association.

Teachers and educational theorists have some control over the sorting-out of English written vocabulary into habit-forming words and analogy-following words. It would obviously be desirable to do this sorting in such a way that the habits established would be analogically extendable with a minimum of error. Thus if there were a good number of important English words in which the initial letter B did not correlate with a pronunciation with initial /b/, the B-to-/b/ habit would be a less reliable one.

One might suspect that the best sorting would be one in which (1) habit-forming words would be those that are regular by the optimum rules (defined above); (2) absolutely essential common words that are irregular by the optimum rules would be taught in some screening context that would forestall or minimize the child's tendency to extend to new material the specific spelling-to-sound correlations in them. That is, we assume that in addition to forming certain useful spelling-to-sound habits, a child can learn to recognize certain words simply as whole units. (Some study suggests that he does this first, and only later comes to associate parts of written words with parts of the pronunciation. But this is a matter of finer-scale detail and need not concern us here.)

GENERAL STRUCTURE OF GRAPHIC MONOSYLLABLES

Each graphic monosyllable includes a vowel spelling. Any letter or letters to the left of a vowel spelling is an initial consonant spelling. Any letter or letters to the right of a vowel spelling is a final consonant spelling.

INITIAL CONSONANT SPELLINGS

The initial consonant spellings that occur in graphic mono-syllables are listed below. All those that have a single regular pronunciation are followed by an indication of that pronunciation. Those for which there are complications needing discussion are merely listed, and then repeated after the full list with the necessary discussion.

B /b/	BL /bl/	BR /br/			
C	CL /kl/	CR /kr/	CH / ^v c/	CHR /kr/	
D /d/		DR /dr/	DW / ^v d/	DH /d/	
F /f/	FL /fl/	FR /fr/			
G	GL /gl/	GR /gr/	GH /g/	GN /n/	GW /gw/
H /h/	J / ^v j/				
K /k/	KL /kl/	KR /kr/	KH /k/	KN /kn/	
L /l/	M /m/	N /n/			
P /p/	PL /pl/	PR /pr/	PH /f/	PHL /fl/	PS /s/
QU /kw/					
R /r/			RH /r/		
S /s/	SL /sl /		SH / ^v s/	SHR / ^v sr/	SW /sw
	SC . .	SCH	SCHW	SCR /skr/	
	SK /sk/				
	SM /sm/				
	SN /sn/				
	SP /sp/	SPL /spl/	SPR /spr/		
	SQU /skw/				
	ST /st/	STR /str/			
T /t/	TR /tr/	TW /tw/			
	TH /θ/	THR /θr/	THW /θw/		
V /v/	W /w/	WH /hw/	WR /r/		
X /z/					
Y /y/					
Z /z/					

Special comments:

C before E, I, Y is regularly /s/; before A, O, U regularly /k/.

G before E, I, Y, except before IR, is regularly /j/; before IR and before A, O, U regularly /g/.

SC before E, I, Y is regularly /s/; before A, O, U regularly /sk/.

SCH is rare, and no single pronunciation is predominant enough to call regular.

SCHW is confined to one word, so that no rule is useful.

GRAPHIC MONOSYLLABLES WITH IRREGULARITIES IN INITIAL CONSONANTISM

CH is /k/ in CHASM CHI CHORD CHOIR CHAOS; /ʃ/ in CHEF.

G is /g/ in GELD GELT GET GIMP GILT GIG GIFT GEAR.

H is silent in HEIR HOUR.

TH is /ð/ in THAN THAT THE THEM THEN THIS THY THEY THEE THEIR THEIRS THOU THOUGH THUS.

WH is /h/ in WHO WHOM; /w/ in WHARF.

TW and SW have silent W in TWO TWOES TWOED SWORD.

QU is /k/ in QUAY.

SCH is /sk/ in SCHOOL; /s/ in SCHISM; /ʃ/ in SCHIST.

SCHW is /ʃw/ in SCHWA.

Rare initials: CZ is /z/ in CZAR; TS and TZ are /ts/ or /z/ in TSAR TZAR.

FINAL CONSONANT SPELLINGS (a few rare ones are omitted from list)

B /b/	BS /bz/	BB /b/	BBS /bz/	BT /t/	BTS /ts/
C /k/	CS /ks/	CH /ç/	CK /k/	CKS /ks/	
			CT /kt/	CTS /kts/	
D /d/	DS /dz/	DD /d/	DDS /dz/	DST /dst/	DSTS /dsts/
		DPH /dθ/	DTHS /dθs/		
F /f/	FS /fs/	FF /f/	FFS /fs/	FT /ft/	FTS /fts/
		FTH /fθ/	FTHS /fθs/		
G /g/	GS /gz/	GG /g/	GGs /gz/	GH /g/	GHS /gz/
		GM /m/	GMS /mz/	GN /n/	GNS /nz/

K /k/	KS /ks/	KH /k/	KHS /ks/		
L /l/	LS /lz/	LB /lb/	LBS /lbz/	LC /lk/	LCS /lks/
		LCH /lč/		LCT /lkt/	LCTS /lcts/
		LD /ld/	LDS /ldz/	LDST /ldst/	
		LF /lf/	LFS /lfs/	LFT /lft/	LFTS /lfts/
		LFTH /lfe/	LFTHS /lfeθs/		
		LK /lk/	LKS /lks/		
		LL /l/	LLS /lz/		
		LM /lm/	LMS /lmz/	LN /ln/	LNS /lnz/
		LP /lp/	LPS /lps/	LPH /lf/	IPHS /lfs/
		LSH /lš/		LT /lt/	LTS /lts/
		LTH /le/	LTHS /leθs/		
		LTZ /lts/		LX /lks/	
M /m/	MS /mz/	MB /m/	MBS /mz/	MN /m/	MNS /mz/
		MP /mp/	MPS /mps/	MPH /mf/	MPHS /mfs/
		MPT /mpt/	MPTS /mpts/		
		MT /mt/	MTS /mts/	MTH /mθ/	MTHS /mθs/
N /n/	NS /nz/	NC /ŋk/	NCS /ŋks/	NCH /nč/	
		NCT /ŋkt/	NCTS /ŋkts/		
		ND /nd/	NDS /ndz/	NDL /nd+l/	NDLS /nd+lz/
		NG /ŋ/	NGS /ŋz/	NGTH /ŋθ/	NGTHS /ŋθs/
		NK /ŋk/	NKS /ŋks/	MN /n/	NNS /nz/
		NST /nst/		NT /nt/	NTS /nts/
		NTH /nθ/	NTHS /nθs/		
		NTZ /nts/		NX /nks/	
P /p/	PS /ps/	PH /f/	PHS /fs/	PP /p/	PPS /ps/
		PT /pt/	PTS /pts/	PTH /pθ/	PTHs /pθs/
R, RH only after vowel spellings ending with R: add nothing					
RS, RHS ditto: add /z/.					
S /z/	SCH /š/	SH /s/	SK /sk/	SKS /sks/	
	SM /z+m/	SMS /z+mz/	SP /sp/	SPS /sps/	
	SS /s/		ST /st/	STS /sts/	
T /t/	TS /ts/	TCH /č/	TH /θ/	THS /θs/	
		THM /ð+m/	THMS /ð+mθ/		
		TT /t/	TTS /ts/	TZ /ts/	
V /v					
X /ks/	ST /kst/	XTS /ksts/	XTH /ksθ/	XTHS /ksθs/	
Z /z/	ZZ /z/				

GRAPHIC MONOSYLLABLES WITH IRREGULARITIES IN FINAL CONSONANTISM

S is /s/ in: GAS YES CHRIS KRIS SIS US BUS PUS PLUS THUS BIAS
DAIS CHAOS.

L as first letter in final consonant spellings is silent in:
AIMS BALM CALM PALM PSALM QUALM CALF HALF GOLF BALK CALK CHALK STALK
TALK WALK BAULK CAULK FOLK YOLK COULD(ST) WOULD(ST) SHOULD(ST).

CHT is /t/ in YACHT.

TH is /ð/ in SMOOTH; /ð/ or /θ/ in WITH.

NCH is /ŋk/ in CONCH.

F is /v/ in OF.

Very rare finals (in addition to CHT as above) are: RH in MYRRH;
NDL in DIRNDL; SCH in BORSCH; CORPS (PS silent in singular, /s/ in
plural); EIGHTH where TH is /tθ/ for most speakers; COUP with silent P.

HOMOGRAPHS

Certain graphic monosyllables represent two words with two pro-
nunciations. No rules can render both pronunciations regular.

WAFt /æ/ or /a/ (same word, in effect, with alternative
pronunciations).

BASS /æ/ or /ey/.

THE /ð/ or /iy/, depending on pronunciation of following word.

WIND /i/ or /ay/.

WITH with final /ð/ or /θ/, free choice for single word.

SIS /sis/ or /siyz/ (latter plural of SI).

DO /duw/ or /dow/.

POLL /a/ (as in POLL PARROT), /ow/.

ROW LOW BOW MOW SOW with /aw/ or /ow/.

READ LEAD with /iy/ or /e/.

TEAR with /ir/ or /er/.

DOES /dæz/ or /dowz/ (latter plural of DOE).

HOOD with /u/ (garment); with /u/ or /uw/ (gangster)

WOUND /uw/ or /aw/

CORPS /kor/ singular, /korz/ plural

VOWEL SPELLINGS

Any occurrence of a vowel letter is part of a vowel spelling, except that U when immediately preceded by Q is part of an initial consonant spelling. There are two vowel spellings that do not involve a vowel letter: Y and YR. When Y is initial in a word it is not part of a vowel spelling. With these restrictions, the vowel spellings in graphic monosyllables are the following:

A	E	I	O	U	Y
AH			OH		
		IGH			
AR	ER	IR	OR	UR	YR
AW	EW		OW		
AY	EY		OY	UY	
	EA	IA	OA	UA	YA
			OAH		
	EAR		OAR	UAR	
AE	EE	IE	OE	UE	YE
					YEI
	EER	IER	OER	UER	YER
		IEW			
AI	EI	II	OI	UI	YI
AIR	EIR		OIR		
AEGH	EIGH				
AO	EO	IO	OO	UO	
				UOW	
			OOH		
			OOR		
AU	EU	IU	OU		
AUGH			OUGH		
			OUR		

Of these, the following occur only in graphic monosyllables that are also monosyllables in pronunciation:

A E I O U Y AH OH IGH AR ER IR OR UR YR AW EW OW AY EY OY
UY OAR AE EE YE EER IEW AIR EIR OIR AIGH EIGH OO UOY OOH
OOR AU EU OU AUGH OUGH OUR.

The following occur only in graphic monosyllables that are pronounced in two syllables:

IA YA OAH YEI OER UER YER II YI AO EO IO UO IU.

The sorting-out of regular and irregular, or the evidence for such sorting out, is presented on the following pages, where appear complete lists of the words with the various spellings. An entry ending in a consonant

other than S, such as CAB, does double duty for itself and for the longer form with "inflectional S" after it: CABS. "Inflectional S" may make a noun plural or the third person singular present tense of a verb: for spelling-to-pronunciation it does not matter which, so long as the pronunciation is the same.

VOWEL SPELLING: A.

Final: /a/. BRA FA HA LA MA PA SPA SCHWA

Before final S:

/a/: BRAS FAS LAS MAS PAS SPAS SCHWAS

/æ/: AS HAS. /æ/, final S irregularly /s/: GAS

Before L, final or followed by further consonant letters:

/æ/. GAL SAL PAL ALB TALC SHALL SHALT ALPS
SCALP PALP CALX

/a/, L silent: ALMS BALM CALM PALM PSALM QUALM

/æ/, L silent: CALF HALF

/ɔ/: SCALD BALD ALL BALL CALL FALL GALL HALL MALL PALL SMALL
SQUALL STALL TALL THRALL WALL HALT MALT SALT WALTZ

/ɔ/, L (usually) silent: BALK CALK CHALK STALK TALK WALK

[Pronunciation predictable by rather complex rules, in terms of the final consonant spelling, save only for SHALL SHALT]

After W, WH, SW, THW, QU, SQU (and not followed by L):

/æ/: WAC QUACK THWACK WHACK SWAG WAG SWAM SWANG TWANG
WHANG SWANK WAX

/a/ or /ɔ/ (depending largely on dialect): SQUAB SWAB QUAD SQUAD

WAD QUAFF SWAMP WAN SWAN WAND WANT
SWAP WAS QUASH SQUASH SWASH WASH WASP SQUAT
SWAT WHAT SWATCH WATCH SWATH WATT

/æ/ or /a/ (or /ɔ/): WAFT

[Thus: /æ/ always before spellings that represent dorso-velar stops or nasal (C, CK, G, NG, NK, X); /a/ or /ɔ/ otherwise save only in SWAM and the variable WAFT]

Isolated irregulars: homographs BASS /bʌs/ and /beys/; YACHT with /a/ and silent CH.

Otherwise: /ə/. Irregular initial consonant spellings in CHASM, THAN, THAT. Words with no irregularities follow.

BABS BLAB CAB CRAB DAB DRAB GAB GRAB JAB NAB SCAB SLAB STAB TAB

MAC SAC

BACK BLACK CLACK CRACK HACK JACK KNACK LACK PACK RACK SACK SHACK SLACK
SMACK SNACK STACK TACK TRACK WRACK

ACT BRACK FACT TACT TRACT

AD BAD BRAD CAD CLAD DAD FAD GAD GLAD HAD LAD MAD PAD SAD SHAD SCADS

ADD

CHAFF GAFF STAFF

AFT CRAFT DAFT DRAFT GRAFT HAFT RAFT SHAFT TAFT

BAG BRAG CRAG DRAG FAG FLAG GAG HAG LAG NAG RAG SAG SCRAG SHAG SLAG SNAG
STAG TAG

FLAK YAK

AM CAM CLAM CRAM DAM DRAM GRAM HAM JAM LAM PRAM RAM SCRAM SHAM SLAM
TAM TRAM YAM

JAMB LAMB

DAMN

CAMP CHAMP CLAMP CRAMP DAMP LAMP RAMP SCAMP STAMP TAMP TRAMP VAMP

AN BAN BRAN CAN CLAN FAN MAN PAN PLAN RAN SCAN SPAN TAN VAN

BLANCH BRANCH RANCH

AND BAND BLAND BRAND GLAND GRAND HAND LAND SAND STAND STRAND

BANG CLANG FANG GANG HANG RANG SANG SLANG SPRANG TANG

BANK BLANK CLANK CRANK DANK DRANK FLANK FRANK HANK LANK PLANK PRANK RANK
SANK SHANK SHRANK SLANK SPANK TANK THANK YANK

MANN BANNS

CANST

ANT BRANT CANT CHANT GRANT PANT PLANT RANT SCANT SLANT

MANX

CAP CHAP CLAP CRAPS FLAP GAP HAP LAP MAP NAP PAP RAP SAP SCRAP SLAP SNAP
STRAP TAP TRAP WRAP YAP

GRAPH

APT RAPT WRAPT

ASH BASH BRASH CASH CLASH CRASH DASH FLASH GASH GNASH HASH LASH MASH PLASH
RASH SASH SMASH SPLASH STASH THRASH TRASH

ASK BASK CASK FLASK MASK TASK

PLASM SPASM

ASP CLASP GASP GRASP HASP RASP

ASS BRASS CLASS CRASS GLASS GRASS LASS MASS PASS

BLAST CAST FAST HAST LAST MAST PAST VAST

AT BAT BLAT BRAT CAT CHAT FAT FLAT GAT GNAT HAT MAT PAT RAT SAT SCAT
SKAT SLAT SPAT SPLAT SPRAT TAT VAT

BATCH CATCH HATCH LATCH MATCH PATCH RATCH SCRATCH SNATCH THATCH

BATH HATH LATH MATH PATH SNATH WRATH

AX FLAX LAX TAX

JAZZ RAZZ

VOWEL SPELLING: E.

Final: /iy/: BE HE ME SHE WE YE

/ey/: RE

/iy/: and /ə/: THE.

Before final S:

/e/, with S irregularly /s/: YES

/ey/: RES.

Otherwise: /e/. Initial G irregularly /g/ in GELD GELT GET.

Initial CH irregularly /ʃ/ in CHEF. Initial TH irregularly
/ð/ in THEM THEN.

BLEB NEB PLEBS

EBB

DEBT

SPECS

BECK CHECK DECK FLECK NECK PECK RECK SPECK WRECK

SECT

FED BED BLED BRED LED PLED RED SHED SHRED SLED SPED TED WED ZED

CLEF

CLEFT DEFT EFT HEFT LEFT REFT THEFT WEFT

BEG DREGS KEG LEG PEG

EGG YEGG

PHLEGM

TREK

GEL EL

BELCH SQUELCH

ELD HELD MELD WELD

ELF PELF SELF SHELF

DELFT

TWELFTH

ELK WHELK YELK

BELL CELL DELL DWELL ELL FELL HELL JELL KNELL QUELL SELL SHELL SMELL SNELL
SPELL SWELL TELL WELL YELL

ELM HELM WHELM

HELP KELP WHELP YELP

WELSH

BELT DWELT FELT KNELT MELT SMELT SPELT WELT

EM GEM HEM SHEM STEM

HEMP

TEMPT

DEN EN FEN GLEN HEN KEN LENS MEN PEN SEN TEN WEN WHEN WREN YEN

BENCH BLENCH CLENCH DRENCH QUENCH STENCH TENCH TRENCH WENCH WRENCH

BEND BLEND END FEND LEND SPEND MEND PEND REND SEND TEND TREND VEND
WEND ZEND

LENGTH STRENGTH

BENT BLENT CENT DENT LENT PENT RENT SENT SCENT SPENT TENT TRENT VENT
WENT

TENTH

HEP PEP PREP REP STEP

REPP

CREPT KEPT SLEPT STEPT SWEPT WEPT

DEPTH

FLESH FRESH MESH THRESH

DESK

BLESS CHESS CRESS DRESS FESS JESS LESS MESS PRESS STRESS TRESS

BEST BLEST CHEST CREST DREST HEST JEST LEST NEST PEST QUEST REST TEST VEST
WEST WREST ZEST

BET FRET JET LET MET NET PET RET SET VET WET WHET YET

FETCH KETCH RETCH SKETCH STRETCH VETCH WRETCH

SETH

METZ

REV

EX FLEX HEX LEX SEX VEX

NEXT SEXT TEXT

VOWEL SPELLING: I.

Final: /ay/: PI PHI PSI XI CHI I

/iy/: MI SI SKI TI

Before final S:

/ay/: PIS PHIS PSIS XIS CHIS

/iy/: MIS SIS SKIS TIS

/i/: CHRIS KRIS SIS (S irregular); IS HIS

Before GN, LD, NTH: /ay/: SIGN CHILD MILD WILD NINTH

Before MB: /ay/ CLIMB but /i/ LIMB

Before ND: /ay/ BIND BLIND FIND GRIND HIND KIND MIND RIND

/ay/ and /i/ (homographs) WIND

Before NT: /ay/ PINT; otherwise /i/: DINT FLINT CLINT HINT LINT MINT PRINT
QUINT SPLINT SPRINT SQUINT STINT TINT

Otherwise: /i/. Initial G irregularly /g/ in GIMP GILT GIG GIFT.

Initial SCH irregularly /s/ in SCHISM, irregularly /ʃ/ in SCHIST.

Initial TH irregularly /ð/ in THIS. Final TH optionally /ə/ or /ð/
in WITH.

BIB CRIB FIB GLIB JIB NIB RIB SIB SQUIB

SIC TIC

RICH WHICH

BRICK CHICK CLICK CRICK FLICK HICK KICK LICK NICK PICK PRICK QUICK
RICK SICK SLICK SNICK STICK THICK TICK TRICK WICK WRICK

STRICT

BID DID FID GRID HID KID LID MID QUID RID SLID SKID SQUID

DIDST MIDST

WIDTH

IF

BIFF CLIFF MIFF SKIFF SNIFF STIFF TIFF WHIFF

DRIFT LIFT RIFT SHIFT SHRIFT SIFT SWIFT THRIFT

FIFTH

BIG BRIG DIG FIG GRIG JIG PIG PRIG RIG SPRIG SWIG TRIG TWIG WHIG WIG

MIL NIL TIL

FILCH MILCH

ILK BILK MILK SILK

BILL BRILL CHILL DILL DRILL FILL FRILL GRILL HILL ILL JILL KILL MILL PILL
QUILL RILL SHRILL SILL SKILL SPILL SQUILL STILL SWILL THILL THRILL
TILL TRILL TWILL WILL

FILM

KILN

HILT JILT KILT LILT MILT QUILT SILT SPILT STILT TILT WILT

FILTH TILTH

BRIM DIM GLIM GRIM HIM PRIM RIM SCRIM SHIM SKIM SLIM SWIM TRIM VIM WHIM

LIMN

BLIMP CRIMP IMP LIMP PIMP PRIMP SCRIMP SHRIMP SKIMP SIMP

BIN CHIN DIN FIN GIN GRIN IN KIN PIN SHIN SIN SKIN SPIN THIN TIN TWIN WIN

ZINC

CHINCH CINCH CLINCH FINCH FLINCH INCH PINCH WINCH

TINCT

BRING CLING DING FLING KING LING PING RING SING SLING SPRING STING STRING
STRING SWING THING TING WING WRING ZING

BLINK BRINK CHINK CLINK DRINK FINK INK KINK LINK MINK PINK PRINK RINK
SHRINK SINK SLINK STINK THINK WINK

INN JINN

PLINTH

CHINTZ

JINX MINX SPHINX

BLIP CHIP CLIP DIP DRIP FLIP GRIP HIP KIP LIP NIP PIP QUIP RIP SCRIP SHIP
SIP SKIP SLIP SNIP STRIP THRIPS TIP TRIP WHIP YIP ZIP

DIPT DRIPT SCRIPT SLIPT

DISH FISH PISH SWISH WHISH WISH

BRISK DISK FRISK RISK WHISK

ISM CHRISM PRISM

CRISP LISP WISP

BLISS HISS KISS MISS SWISS

GIST HIST LIST MIST TWIST WHIST WRIST

BIT CHIT FIT FLIT GRIT HIT IT KIT KNIT LIT NIT PIT QUIT SIT SKIT SLIT
SMIT SPIT SPLIT SPRIT TIT TWIT WHIT WIT WRIT

BITCH DITCH FLITCH HITCH ITCH PITCH QUITCH SNITCH STITCH SWITCH TWITCH WITCH

PITH SITH KITH SMITH

BITT MITT BLITZ SPITZ FIX MIX NIX SIX FIXT :SIXTH

FIZ FRIZ QUIZ WHIZ FIZZ

VOWEL SPELLING: O

Final: /ow/: FRO GO HO JO LO NO PRO RHO SO WO.

/UW/: TO. Initial irregular in WHO TWO.

/ow/ or /uw/ (homographs):DO.

Before final S:

/ow/: NOS PROS RHOS.

Before L plus consonant:

/ow/: OLD BOLD COLD FOLD GOLD HOLD MOLD SOLD SCOLD TOLD WOLD BOLL
DROLL KNOLL ROLL SCROLL STROLL TOLL TROLL HOLM HOLF BOLT COLT
JOLT MOLT SMOLT VOLT

/ow/ or /a/: POLL

/ow/, L silent:OLPOLL YOLK

/u/: WOLF

/a/: DOLL MOLL LOLL. /a/ or /ɔ/, usually silent: GOLF

Before M or M plus consonant:

/ow/: COMB

/uw/: TOMB WOMB. Initial irregular in WHOM

/ə/: FROM

/a/: PROM PROMPT BOMB POMP ROMP

Before H or N plus consonant:

/ə/: SON TON WON MONK FRONT MONTH

/a/ (or /ʌ/), depending largely on dialect): ON CON DON YON BLOND
BOND FOND FROND POND YOND FONT GONG LONG PRONG SONG STRONG
TONG THONG THrong WRONG HONK WONT

Before S plus consonant:

-SH, -SK: /a/ (or /ʌ/), depending on dialect: BOSH JOSH SLOSH BOSK

-SS: /ow/ in GROSS, otherwise /a/ (or /ʌ/): BOSS CROSS DROSS FLOSS.
GLOSS LOSS MOSS ROSS TOSS JOSS

-ST: /ə/ in DOST; /ow/ in GHOST HOST MOST POST; otherwise /a/ (or /ʌ/): COST FROST LOST

Before TH: /ə/ in DOTH; /ow/ in BOTH, LOTH; otherwise /a/ (or /ʌ/):
BROTH CLOTH FROTH MOTH SLOTH TROTH WROTH

Otherwise: /a/ or /ʌ/, depending largely on dialect. Some dialects have only one pronunciation in all words. Some have both, but for these dialects it depends almost always on the following consonantism: e.g., /a/ before T (pronounced /t/), but /ʌ/ before FT (pronounced ft), and the like.

BOB BLOB COB FOB GOB HOB JOB KNOB LOB MOB NOB ROB SOB SLOB SNOB THROB

BLOC BOCK BLOCK CHOCK CLOCK COCK CROCK DOCK FLOCK FROCK HOCK KNOCK LOCK
MOCK NOCK ROCK SOCK SHOCK SMOCK STOCK

CLOD COD GOD HOD NOD PLOD POD PROD ROD SCROD SHOD SOD TROD ODD

OFF DOFF SCOFF OFT CROFT LOFT SOFT

BOG COG CLOG DOG FOG FLOG FROG GROG HOG JOG LOG SLOG SMOG TOG

CHOP COP CROP DROP FLOP POP HOP LOP MOP PLOP POP PROP SHOP SLOP SOP STOP
STROP TOP DROPT STOPT

BLOT BOT CLOT COT DOT GOT HOT JOT KNOT LOT PLOT POT ROT SOT SCOT
SHOT SLOT SPOT TOT TROT WOT BLOTCH BOTCH CROTCH NOTCH SCOTCH
SPLOTCH SCOTT

OX BOX COX FOX PHLOX POX VOX

VOWEL SPELLING: U

Final: /(y)uw/: FLU GNU MU NU THRU.

Before final S: /(y)uw/: FLUS GNUS MUS NUS; but /ə/ with final S irregularly
/s/ in US BUS PUS PLUS THUS

Before LL: /u/ in BULL FULL PULL; otherwise /ə/: CULL DULL GULL HULL
LULL MULL NULL SCULL SKULL TRULL

Before SH: /u/ in BUSH PUSH; otherwise /ə/: BLUSH BRUSH CRUSH FLUSH GUSH
HUSH LUSH MUSH PLUSH RUSH SLUSH TUSH THRUSH

Before SS: /u/ in PUSS; otherwise /ə/: BUSS CUSS FUSS MUSS TRUSS

Before T: /u/ in PUT; otherwise /ə/: BUT CUT GUT GLUT HUT JUT NUT RUT
SCUT SHUT SLUT SMUT STRUT TUT

Otherwise: /ə/: CUB CHUB CLUB DUB DRUB GRUB HUB NUB SUB SCRUB SHRUB SNUB
STUB RUB TUB

MUCH SUCH BUCK CHUCK CLUCK DUCK LUCK MUCK PLUCK RUCK SUCK SHUCK STUCK
STRUCK TUCK TRUCK DUCT

BUD CUD DUD MUD SCUD SPUD STUD THUD SUDS

BUFF BLUFF CUFF DUFF FLUFF GUFF HUFF LUFF MUFF PUFF RUFF SCUFF SCRUFF
SNUFF STUFF TUFF TUFT

BUG CHUG DUG DRUG HUG JUG LUG MUG PUG PLUG RUG SHRUG SLUG SMUG SNUG
TUG THUG

BULB GULCH MULCH MULCT GULF BULK HULK SULK SKULK CULM
GULP PULP CULT

BUM CHUM DRUM GUM GLUM HUM MUM PLUM RUM SUM SCUM SLUM STRUM SWUM
THRUM CRUMB DUMB NUMB PLUMB RHUMB THUMB BUMP CHUMP
CLUMP CRUMP DUMP FRUMP HUMP JUMP LUMP PUMP PLUMP RUMP SUMP
SLUMP STUMP THUMP TRUMP MUMPS

BUN DUN FUN GUN HUN NUN PUN RUN SUN SPUN STUN SHUN TUN BUNCH
CRUNCH HUNCH LUNCH MUNCH PUNCH SCRUNCH FUND BUNG CLUNG
DUNG FLUNG HUNG LUNG RUNG SUNG SLUNG SPRUNG STUNG STRUNG SWUNG
WRUNG BUNK CHUNK DUNK DRUNK FUNK FLUNK HUNK JUNK PUNK
PLUNK SUNK SHRUNK SKUNK SLUNK SPUNK STUNK TRUNK BUNT
BLUNT BRUNT GRUNT HUNT PUNT RUNT SHUNT STUNT

UP CUP PUP SUP SCUP

CUSK DUSK HUSK RUSK MUSK TUSK CUSP BUST CRUST DUST GUST JUST
LUST MUST RUST TRUST THRUST

CLUTCH CRUTCH HUTCH SCUTCH SMUTCH BUTT MUTT PUTT

CRUX FLUX TUX

BUZZ FUZZ

VOWEL SPELLING: Y.

Final: /ay/: BY CRY DRY FLY FRY MY PLY PRY SHY SLY SKY SPY SPRY STY TRY
WHY WRY. Initial TH irregularly /ð/: THY

Followed by anything: /i/: SYLPH GYM HYMN LYMPH NYMPH LYNCH LYNX GYP
CRYPT CYST TRYST MYTH PYX STYX. Final THM is
/ðtm/: RHYTHM.

VOWEL SPELLING: AH.

Always /a/: AH BAH HAH PAH RAH SHAH YAH

VOWEL SPELLING: OH

/ow/: OH, OHM. /a/: JOHN

VOWEL SPELLING: IGH

Always /ay/: HIGH NIGH SIGH THIGH BIGHT BRIGHT BLIGHT DIGHT FLIGHT
FRIGHT HIGHT KNIGHT LIGHT MIGHT NIGHT PLIGHT RIGHT
SIGHT SLIGHT TIGHT WIGHT WRIGHT

VOWEL SPELLING: AR

After W, SW, DW, WH, QU, THW, /or/ or /ɔr/: WAR SWORD WARD DWARF WHARF
WARM SWARM WARMTH WARM WARP QUART SWART THWART WART
SWARTH QUARTZ

Otherwise /ar/: BAR CAR CHAR FAR GAR JAR MAR PAR SCAR SPAR STAR TAR
CZAR TSAR TZAR

BARB

ARC ARCH LARCH MARCH PARCH STARCH

BARD CARD CHARD HARD LARD NARD PARD SHARD YARD

SCARF

ARK BARK DARK LARK MARK PARK SHARK SPARK STARK

CARL GNARL JARL MARL SNARL

ARM BARM CHARM FARM HARM

BARN DARN TARN YARN

CARP HARP SCARP SHARP

PARR

MARS HARSH MARSH

ART CART CHART DART HART MART PART SMART START TART

MARX

VOWEL SPELLINGS: ER IR UR YR

All always /ər/.

HER PER FIR SIR STIR WHIR BUR BLUR CUR FUR PUR SLUR SPUR

ERR CHIRR SHIRR WHIRR BURR CHURR PURR

MYRRH

HERB KERB SERB VERB BLURB CURB

PERCH BIRCH SMIRCH CHURCH LURCH

HERD SHERD BIRD GIRD THIRD CURD SURD

KERF SERF SURF SCURF TURF

ERG BERG BURG BURGH

CLERK JERK PERK DIRK KIRK MIRK QUIRK SHIRK SMIRK LURK MURK TURK
MERL GIRL SKIRL SWIRL TWIRL WHIRL BURL CURL CHURL HURL FURL KNURL PURL
GERM PERM SPERM TERM FIRM SQUIRM
FERN KERN QUERN STERN TERN URN BURN CHURN SPURN TURN DIRNDL
BURNT

CHIRP STIRPS
HERS ERST FIRST THIRST BURST CURST DURST
PERT WERT DIRT FLIRT GIRT QUIRT SHIRT SKIRT SPIRT SQUIRT BLURT CURT HURT
SPURT BERTH PERTH BIRTH FIRTH GIRTH MIRTH
[N.B.: C is regularly /j/ before ER, but always /g/ before IR.]

VOWEL SPELLING: OR

After W: /ɔr/ in WORD WORK WORLD WORM WORST WORTH; otherwise /or/: WORN WORT.
Otherwise /or/: OR FOR NOR ORB PORCH SCORCH TORCH CORD CHORD
FORD LORD SWORD CORK FORK PORK STORK YORK WHORL CORM
DORM FORM NORM STORM BORN CORN HORN LORN MORN SCORN SHORN SWORN
TORN THORN THORP CORPS BORSCH PORT SORT SHORT
SNORT SPORT TORT FORTH NORTH

VOWEL SPELLING: AW

Always /ɔ/.
BRAW CAW CLAW CRAW DAW DRAW FLAW GNAW HAW JAW LAW MAW PAW RAW SAW
SHAW SLAW SQUAW STRAW TAW THAW YAW
BAWD
HAWK SQUAWK AWL BAWL BRAWL CRAWL DRAWL PAWL SHAWL SPRAWL TRAWL YAWL
SHAWN AWN BRAWN DAWN DRAWN FAWN LAWN PAWN PRAWN SAWN SPAWN
YAWN YAWP

VOWEL SPELLING: EW

/ow/ in: SEW SEWS SEWN
Otherwise / (y)uw/: BLEW BREW CHEW CLEW CREW DEW DREW FEW FLEW HEW JEW
KNEW MEW NEW PEW SCREW SHREW SKEY SLEW SPEW STEW STREW THREW YEW
LEWD SHREWD MEWL HEWN STREWN THEWS NEWT

VOWEL SPELLING: OW

/ow/ in:	/aw/ in:
BLOW CROW FLOW GLOW GROW KNOW	BROW COW CHOW HOW NOW PLOW
SHOW SLOW SNOW STOW STROW	PROW SCOW TROW VOW
TOW THROW	
BOWL	CROWD
	OWL COWL FOWL GROWL HOWL JOWL
	PROWL SCOWL YOWL
OWN BLOWN FLOWN SHOWN MOWN	BROWN CLOWN CROWN DOWN DROWN
SOWN STROWN THROWN	FROWN GOWN TOWN

GROWTH

Homographs /ow/ and /aw/: ROW LOW BOW NOW SOW.

VOWEL SPELLING: AY

/e/ in SAYS; /iy/ in QUAY.

Otherwise /ey/: BAY BRAY CAY CLAY DAY DRAY FAY FLAY FRAY GAY
GRAY HAY JAY LAY MAY NAY PAY PLAY PRAY RAY SAY SHAY SLAY
SPAY SPLAY SPRAY STAY STRAY SWAY TRAY WAY

VOWEL SPELLING: EY

/iy/: KEY.

/ey/: BEY DEY GREY HEY PREY TREY WHEY THEY.

VOWEL SPELLING: OY

/oy/: BOY COY CLOY JOY SOY TOY TROY.

VOWEL SPELLING: UY

/ay/: BUY GUY

VOWEL SPELLING: EA

Before D: /e/ in BREAD DEAD DREAD HEAD SPREAD STEAD THREAD TREAD
BREADTH; /iy/ in BEAD KNEAD MEAD PLEAD. Homographs with /e/
and /iy/: READ LEAD.

Before F: /e/ in DEAF; /iy/ in LEAF SHEAF.

Before L: /e/ in REALM DEALT HEALTH STEALTH WEALTH; /iy/ in
DEAL HEAL LEAL MEAL PEAL REAL SEAL SQUEAL STEAL TEAL VEAL
WEAL WHEAL ZEAL WEALD.

Before K: /ey/ in STEAK BREAK; /iy/ in CREAK BLEAK FREAK LEAK
PEAK SNEAK SPEAK SQUEAK STREAK TEAK TWEAK WEAK WREAK

Before MT, NT, PT: /e/ in DREAMT LEANT MEANT LEAPT.

Before ST: /e/ in BREAST; /iy/ in EAST BEAST FEAST LEAST YEAST.

Before T: /e/ in SWEAT; /ey/ in GREAT; /iy/ in EAT BEAT BLEAT
CHEAT CLEAT FEAT HEAT MEAT NEAT PEAT PLEAT SEAT TEAT TREAT
WHEAT

Before TH: /e/ in BREATH DEATH; /iy/ in HEATH NEATH SHEATH
WREATH.

Final (or before S): /ey/ in YEA; /iy/ in FLEA LEA PEA PLEA
SEA TEA.

/iyə/ in REACT.

Before N: /iyt/ in PEAN; otherwise /iy/: BEAN CLEAN DEAN
GLEAN LEAN MEAN QUEAN WEAN

Otherwise /iy/: EACH BEACH BLEACH BREACH LEACH PEACH PLEACH PREACH REACH
TEACH BEAM BREAM CREAM DREAM GLEAM REAM SEAM SCREAM STEAM
STREAM TEAM CHEAP HEAP LEAP NEAP REAP LEASH

[Calling /iy/ "regular," anything else "irregular," there are 99 regulars,
29 irregulars, and 2 homographs.]

VOWEL SPELLING: IA

Always two syllables starting with /ay/. /áy+/ in DIAL PLIANT GIANT
PRIAM STRIA VIA PHIAL TRIAL VIAL BIAS. /áyæ/ in LAMB TRIAD.
/áyē/ or /áyē/ in SIAM.

VOWEL SPELLING: IAR

/áyr/ LIAR

VOWEL SPELLING: OA

Final (or before S): /ow/ in WHOA; /ówə/ in MOA PROA STOA BOA GOA.

Before D: /ɔ/ in BROAD; /ow/ in GOAD LOAD ROAD TOAD WOAD

Before B: /ówə/ in MOAB.

Otherwise /ow/: BROACH COACH LOACH POACH ROACH OAF LOAF OAK CLOAK
CROAK SOAK COAL FOAL GOAL SHOAL SKOAL FOAM LOAM ROAM
GROAN LOAN MOAN ROAN SOAP BOAST COAST ROAST TOAST OATS BOAT
BLOAT COAT DOAT FLOAT GOAT GLOAT GROAT MOAT SHOAT STOAT
THROAT OATH LOATH COAX HOAX

VOWEL SPELLING: UA (R)

/(y)úwə/; /(y)úwt/: DUAL SKUA STUART

/a/: GUARD

VOWEL SPELLING: YA

/áyæ/: DYAD DRYAD

VOWEL SPELLING: OAH /ówə/: NOAH

VOWEL SPELLING: EAR

Final (or before S): /er/ in BEAR PEAR SWEAR WEAR; /ir/ in EAR BLEAR
CLEAR DEAR DREAR FEAR GEAR HEAR NEAR REAR SEAR SHEAR SMEAR
SPEAR YEAR; homograph /er/ and /ir/ in TEAR

Before consonants other than S: /ir/ in BEARD; /ar/ in HEART HEARTH;
/iyár/ in REARM; /ɔr/ in SEARCH HEARD EARL PEARL EARN LEARN
YEARN LEARN EARTH DEARTH

VOWEL SPELLING: OAR

/or/ or /ɔr/ (depending on dialect): OAR BOAR HOAR ROAR SOAR BOARD
HOARD

VOWEL SPELLING: AE /ey/: BRAE NAE

VOWEL SPELLING: EE

/iy/: BEE FEE FLEE FREE GEE CHEE GLEE KNEE LEE SEE SKEE SPREE TEE
THREE THEE TREE WEE BEECH BREECH LEECH SCREECH SPEECH
BLEED BREED CREED DEED FEED GREED HEED MEED NEED REED SEED SCREED
SPEED STEED TWEED WEED BEEF REEF CHEEK CLEEK CREEK GREEK
LEEK MEEK PEEK REEK SEEK SLEEK WEEK EEL CREEL FEEL HEEL KEEL
KNEEL PEEL REEL STEEL WHEEL DEEM SEEM TEEM GREEN KEEN
QUEEN SEEN SCREEN SHEEN SPLEEN WEEN TEENS CHEEP CREEP DEEP
JEEP KEEP PEEP SEEP SHEEP SLEEP STEEP SWEEP VEEP WEEP BEET
FEET FLEET GREET MEET SHEET SLEET STREET SWEET TWEET TEETH

VOWEL SPELLING: IE

Final, before S, before D: /ay/: DIE PIE LIE PIE TIE VIE DIES LIES
PIES TIES VIES CRIES SKIES DRIES FRIES TRIES DIED LIED PIED
TIED VIED CRIED DRIED FRIED TRIED

Before consonants, otherwise: /e/ in FRIEND; /áy+/ in DIET QUIET;
/iy/ in BRIEF CHIEF FIEF GRIEF LIEF THIEF SHRIEK SPIEL FIELD
WIELD YIELD LIEN MIEN PRIEST

VOWEL SPELLING: OE

Final: /uw/ in SHOE; /ow/ in DOE FLOE HOE JOE POE SLOE ROE TOE
THROE WOE

Before S: /uw/ in SHOES TWOES; homographs /ow/ and /ə/ in DOES;
/ow/ in DOES FLOES HOES JOES TOES THROES WOES

Before D: /uw/ in SHOED TWOED; /ow/ in HOED TOED JOES

Before consonants, otherwise: /ówt/ in POEM POET; /òwé/ in NOEL;
/úwt/ in DOETH.

VOWEL SPELLING: UE

Final (or before S or D): BLUE CUE CLUE DUE FLUE GLUE HUE RUE SUE
SLUE SPUE TRUE BLUES CUES CLUES DUES FLUES GLUES HUES RUES
SUES SLUES SPUES TRUES BLUED CUED CLUED GLUED HUED RUED SUED
TRUED

Otherwise: /e/ in GUESS GUEST; /úwt/ in GRUEL SUET CRUEL DUEL CRUET;
/úwé/ in SUEZ; /ùwé/ in DUET

VOWEL SPELLING: YE /ay/: DYE LYE DYES LYES DYED

VOWEL SPELLING: YEI /áyi/: DYEING

VOWEL SPELLING: EER

/ir/: BEER CHEER DEER FLEER JEER LEER PEER QUEER SEER SHEER SNEER
STEER VEER

VOWEL SPELLING: IER

/ɪr/ in BIER PIER TIER; /éytr/ in BRIER CRIER DRIER FLIER FRIER PLIERS
PRIER TRIER

VOWEL SPELLING: OER /úwtr/: DOER

VOWEL SPELLING: UER /úwtr/: BLUER CUER GLUER SUER TRUER

VOWEL SPELLING: YER /eytr/: DYER

VOWEL SPELLING: IEW /(y)uw/: VIEW

VOWEL SPELLING: AI:

/e/ in SAID SAITH; /æ/ in PLAID; /éyi/ in LAIC DAIS

Otherwise /ey/: AID BRAID LAID MAID PAID RAID STAIT WAIF AIL BAIL
BRAIL FAIL FLAIL FRAIL HAIL JAIL KAIL MAIL NAIL PAIL QUAIL RAIL
SAIL SNAIL TAIL TRAIL VAIL WAIL AIM CLAIM MAIM BLAIN BRAIN
CHAIN DRAIN FAIN GAIN GRAIN LAIN MAIN PAIN PLAIN RAIN SLAIN
SPRAIN STAIN STRAIN SWAIN TRAIN TWAIN VAIN WAIN FAINT PAINT
QUAINT SAINT TAIT WAIST BAIT GAIT PLAIT TRAIT STRAIT
WAIT FAITH WRAITH

VOWEL SPELLING: EI

/iy/ in CEIL SHEIK (H); /ay/ in STEIN; /íyi/ or /íyt/ in THEISM THEIST
DEISM DEIST BEING. Otherwise /ey/: LEI DEIGN FEIGN REIGN VEIL
REIN SKEIN VEIN FEINT

VOWEL SPELLING: II /íyi/ SKIING

VOWEL SPELLING: OI

/ówi/ GOING; /úwi/ DOING.

/oy/: POI YOICKS VOID COIF COIGN OIL BOIL BROIL COIL FOIL MOIL ROIL
SOIL SPOIL TOIL COIN GROIN JOIN LOIN QUOIN JOINT POINT FOIST HOIST
JOIST MOIST DOIT QUOIT

VOWEL SPELLING: UI

/(y)uw/: FRUIT SUIT BRUIT

/i/: BUILD BUILT GUILD GUILT

/(y)úwi/ or /(y)úwt/: RUIN BRUIN BLUING BLUISH DRUID TRUISM

VOWEL SPELLING: YI

/áyi/: DRYING FLYING FRYING CRYING PLYING PRYING SHYING SPYING
TRYING DYING

VOWEL SPELLING: AIR

/er/: AIR CHAIR FAIR GLAIR HAIR LAIR PAIR STAIR VAIR LAIRD BAIRN CAIRN

VOWEL SPELLING: EIR

/ir/: WEIR WEIRD; /er/: HEIR THEIR THEIRS

VOWEL SPELLING: OIR /óy(ə)r/: COIR; /wáy(ə)r/: CHOIR

VOWEL SPELLING: AIGH /ey/: STRAIGHT

VOWEL SPELLING: EIGH

/ay/: HEIGHT SLEIGHT; /ey/: NEIGH SLEIGH WEIGH EIGHT FREIGHT
WEIGHT EIGHTH

VOWEL SPELLING: AO /éyà/: CHAOS

VOWEL SPELLING: EO /íyà/: EON LEON NEON PEON

VOWEL SPELLING: IO

/áy+/: VIOL ION LION SCION ZION PRIOR; /íyà/ in KIOSK;
/íyòw/ in TRIO

VOWEL SPELLING: OO

Before D: /uw/ in BROOD MOOD ROOD: SNOOD; /u/ in GOOD STOOD WOOD;
/uw/ or /u/ in FOOD HOOD (the latter in two different senses);
/ə/ in BLOOD FLOOD

Before F: /uw/ in PROOF SPOOF; /u/ in WOOF; /uw/ or /u/ in HOOF ROOF.

Before K: /uw/ in SPOOK; /u/ in BOOK BROOK COOK CROOK HOOK LOOK NOOK
ROOK SHOOK TOOK.

Before L: /uw/ in COOL DROOL POOL POOL SCHOOL SPOOL STOOL TOOL;
/u/ in WOOL.

Before M: /uw/ in BOOM BLOOM BROOM DOOM GLOOM GROOM LOOM ZOOM;
/uw/ or /u/ in ROOM

Before P: /uw/ in COOP DROOP LOOP POOP SCOOP SLOOP SNOOP STOOP SWOOP
TROOP WHOOP; /uw/ or /u/ in HOOP

Before T: /uw/ in BOOT COOT HOOT LOOT MOOT SCOOT SHOOT SNOOT;
/u/ in FOOT; /uw/ or /u/ in ROOT SOOT TOOT(S)

Otherwise /uw/: BOO COO LOO MOO SHOO TOO WOO ZOO BOOB BROOCH
HOOSH MOOSH POOSH BOON COON CROON COON LOON MOON NOON SOON
SPOON SWOON BOOST ROOST BOOTH TOOTH SOOTH SMOOTH

VOWEL SPELLING: UO /úwòw/: DUO

VOWEL SPELLING: EOW /yaw/ or /iyáw/: MEOW

VOWEL SPELLING: UOY /oy/ or /uy/: BUOY

VOWEL SPELLING: OOH /uw/: POOH

VOWEL SPELLING: COR

/ur/ in BOOR MOOR POOR SPOOR; /or/ or ɔr/ in DOOR FLOOR

VOWEL SPELLING: AU

/ɔ/ or aw/ in TAU.

Otherwise /ɔ/: DAUB FRAUD LAUD AUK CAUL HAUL MAUL PAUL SAUL AULD
 FAULT VAULT FAUN HAUNCH LAUNCH PAUNCH STAUNCH CHAUNT DAUNT
 FLAUNT GAUNT HAUNT TAUNT VAUNT TAUT NAUTCH; BAULK CAULK
 (silent L)

VOWEL SPELLING EU /ɛw/ : FEUD RHEUM SLEUTH ZEUS

VOWEL SPELLING: IU /ɛyɔ/: TRIUMPH

VOWEL SPELLING: OU

Before CH: /aw/ in OUCH COUGH CROUCH GROUCH POUCH SLOUCH VOUGH;
 /ə/ in TOUCH

Before L: /aw/ in FOUL; /ow/ in SOUL; /uw/ in CHOUL

Before L plus consonant: /ow/ in MOULD MOULT; /u/ with silent L :
 in COULD(ST) WOULD(ST) SHOULD(ST)

Before TH: /aw/ in DROUTH MOUTH SOUTH; /uw/ in YOUTH

Before P: /uw/ CROUP GROUP SOUP COUP (silent P)

Before NG: /ə/ YOUNG

Before ND: /uw/ or /aw/ in WOUND; /aw/ in BOUND FOUND GROUND MOUND
 POUND ROUND SOUND ZOUNDS

Final: /aw/ in THOU; /uw/ in YOU.

Otherwise /aw/: CLOUD LOUD PROUD SHROUD NOUN COUNT FOUNT MOUNT
 JOUST OUT BOUT CLOUT FLOUT GOUT GROUT KNOUT LOUT POUT ROUT
 SCOUT SHOUT SNOUT SPOUT SPROUT STOUT TROUT DOUBT

VOWEL SPELLING: AUGH

/ɔ/ in AUGHT NAUGHT TAUGHT; /æf/ in LAUGH DRAUGHT(S)

VOWEL SPELLING: OUGH

Before T: /ɔ/ in OUGHT BOUGHT BROUGHT NOUGHT SOUGHT THOUGHT WROUGHT;
 /aw/ in DROUGHT.

Final (or before S): /uw/ THROUGH; /ow/ DOUGH THOUGH; /aw/
 BOUGH PLOUGH; /æf/: CHOUGH ROUGH TOUGH; /ɔf/ COUGH TROUGH.

VOWEL SPELLING: OUR /or/: FOUR POUR GOURD MOURN COURT FOURTH; /ur/:

DOUR TOUR YOUR(S); /aw/(+r)/: FLOUR HOUR SOUR SCOUR OUR(S)

A COMPUTER PROGRAM FOR DERIVING SPELLING-TO-SOUND CORRELATIONS¹

Richard L. Venezky

Cornell University

1. INTRODUCTION

In the history of all natural languages, writing systems, if they exist, appeared long after the spoken language. The primary function of a writing system is to create a permanent representation of speech and therefore we expect not only a correlation between the units of the graphemic system and some units of the spoken language, but, at least in the early stages of the development of the written language, a reasonably regular correspondence.

A graphemic symbol is either phonemic or morphemic, depending on whether it is correlated with phonological or meaningful elements of the spoken language. Graphemic systems are either phonemic, morphemic, or a combination of the two. The English graphemic system originated as a phonemic writing system around the 7th century and was derived primarily from the extant Latin alphabet. As the spoken language followed its natural course of sound change through the centuries, the graphemic system changed also, but the directions of change were seldom parallel.

By the time of the last changes in the graphic system in the 17th and 18th centuries, the original direct correlations between graphemes and phonemes had become a twisted mass of irregular relations. In spite of this problem, however, most people today still learn to

1. This is an abbreviated version of a dissertation submitted to Cornell University for the M. A. degree. The complete thesis including the computer program is available from the Cornell University Library.

correlate graphemes and phonemes, that is, learn to read, even though in most cases no overtly statable spelling-to-sound correlations are formed.

1.1. The exact nature of correlations between spelling and sound, however, is at present a concern of several different disciplines. Primary school teachers need this information primarily to teach reading. To facilitate this learning process, we should like the first words that a child learns to pronounce in reading to be those words that have the most regular spelling-to-sound correlations; that is, those that have the greatest transfer value to new words that the child will meet. This requirement holds true for teaching reading to adults and for second language learning, where a knowledge of spelling-to-sound correlations that differ from those in one's native language serves as a short cut towards learning to read and pronounce the new language.

In the development of a reading machine, a machine that will automatically scan written material and read the text aloud, a spelling-to-sound correlation-network serves as a communication link between the optical scanner and the vocal output unit. While this link could conceivably be a large-scale computer which stores the pronunciations of every word in the English language, such a system is not practicable and will not be necessary once all spelling-to-sound correlations for English are known.

The type of information needed for the above-mentioned requirements is at present not available. While numerous studies of spelling-to-sound correlations have been made in the past, none have tabulated the frequencies of occurrences of the all possible correlations and none have produced exhaustive listings of common words with irregular

pronunciations. What is needed is, first, a complete tabulation of every spelling-to-sound correlation that occurs in the words found in an average recognition vocabulary (which we are arbitrarily assuming in this report to be in the vicinity of 20,000 words), and, second, a listing of these words dichotomized into regular and irregular pronunciation groups. The labels "regular and "irregular," will, of course, be designated on the basis of the spelling-to-sound statistics derived in the tabulation process.

1.2. The Reading Research Project at Cornell University, in studies directed toward the designing of efficient reading materials for the primary schools, started compiling this information in 1959. A complete study of graphic monosyllables was done by C. F. Hockett and a similar study of disyllables was carried out by I. Shah. Both of these studies were done completely by hand and required almost two years to complete. It became evident with the conclusion of this work that only a large computer could be expected to complete the task in a reasonable length of time. With the impending purchase of a Control Data 1604 computer by Cornell, the final impetus was given for finding a machine solution to this problem, and the program described here was begun.

The purpose of this paper is to describe the program that has been designed for deriving spelling-to-sound correlations for English. While the motivation for and the primary orientation of this work is toward the improvement of teaching materials for the primary grades, the material derived from this program will be equally applicable to the other two areas mentioned in the introduction.

and phonemes. Graphemes and phonemes will be referred to occasionally as G's and P's. Specific graphemes will be enclosed in brackets, [X] while specific phonemes will be enclosed by diagonal lines, /X/.

1.3 THE PROBLEM

Programming a computer to derive spelling-to-sound correlations involves two distinct processes. In the first, correlations are derived for each dictionary word and stored. In the second, tabulations are made of the correlations formed in the first part, based upon position in the word and possible environmental influences. Since the second part involves standard tabulation procedures and presents no problems unique to the derivations of spelling-to-sound correlations, it will be discussed only briefly in section 11. The first part of this program, however, is more complex and will be discussed in detail in this section.

1.4. For purposes of this discussion, we will consider a computer to be a device with an unlimited storage space that can add, subtract, multiply, and divide machine words, and can store these words, accept new machine words as input, and transmit machine words out.

A grapheme is any of the 26 letters of the English alphabet. A dictionary word, or grapheme word is a sequence of graphemes that appears as a single entry in a dictionary. A phoneme is any of the forms used in the Thorndike Senior Century Dictionary to indicate the pronunciation of a grapheme word. These forms along with their approximate Smith-Trager phonemic equivalents are given in Appendix III. Some phonemes are represented in print by the same forms used to represent graphemes. A machine word is the contents of a storage location in the computer. Each machine word in the computer used for this program can hold eight graphemes, or eight phonemes, or any combination of eight graphemes and phonemes. Graphemes and phonemes will be referred to occasionally as G's and P's. Specific graphemes will be enclosed in brackets, [X] while specific phonemes will be enclosed by diagonal lines, /X/.

1.5. Given a machine with these characteristics, the problem is to instruct it to derive the correct spelling-to-sound correlations for each grapheme-phoneme pair fed into it. To a sapient human, this is a trivial problem, but to the Machine-Which-Is-Not-A-Brain, capable only of the functions mentioned earlier, this represents a formidable task.

What we want to do is to feed into the computer a sequence of graphemes and a corresponding sequence of phonemes and have the computer decide which phonemes correlate with which graphemes. If the English alphabet were an ideally phonetic alphabet, so that each grapheme had one and only one corresponding phoneme, then this would be a trivial matching problem. Unfortunately, this is not the case in English. Given a sequence of only three graphemes which correspond to a string of three phonemes, any of the correlation patterns shown in Figure 1 and undoubtedly many more are possible. As the number of graphemes increases linearly, furthermore, the number of possible correspondence patterns increases in an irregular manner. This problem is illustrated further in Figure 2 where a number of possible correspondence patterns for one-, two-, and three-grapheme sequences are shown.

1.6. What we have to begin with is a finite set of elements, called graphemes, which can be connected in strings, a string being any finite, linear sequence of these elements. Well-formed strings are formed by applying construction rules to the set of graphemes, the construction rules being an open ended set of rules that generate from the graphemes all and only those words that occur in the dictionary at that time.

Corresponding to every well-formed string of graphemes there exists a string of phonemes into which the graphemes can be mapped.

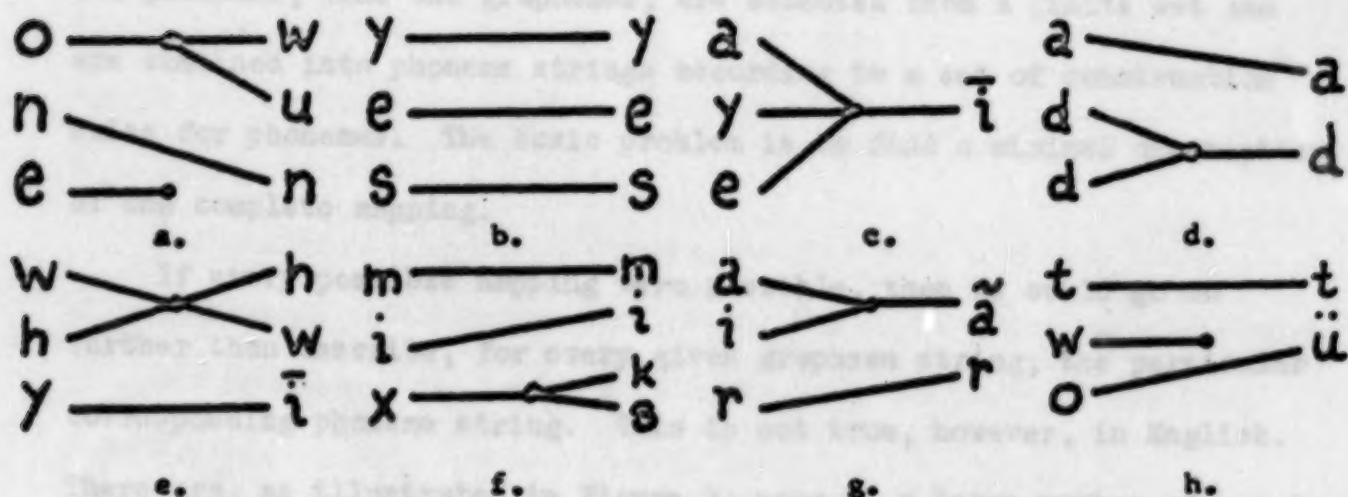


FIG. 1. MAPPINGS OF THREE GRAPHEMES INTO THREE PHONEMES

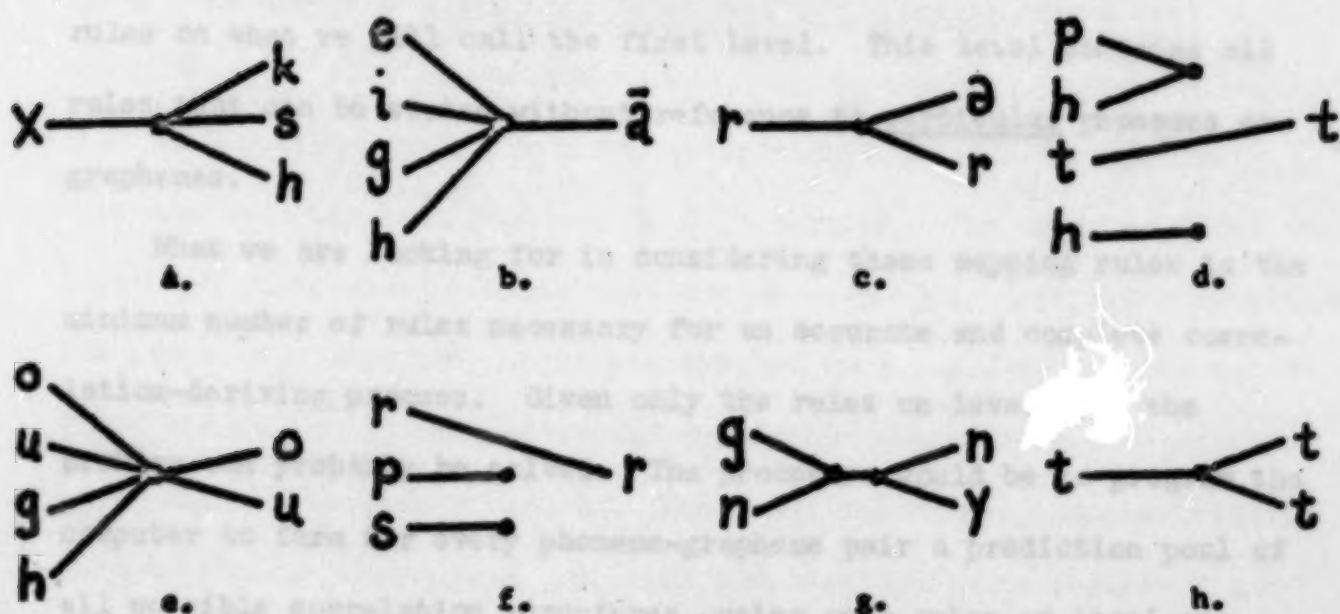


FIG. 2. IRREGULAR MAPPINGS

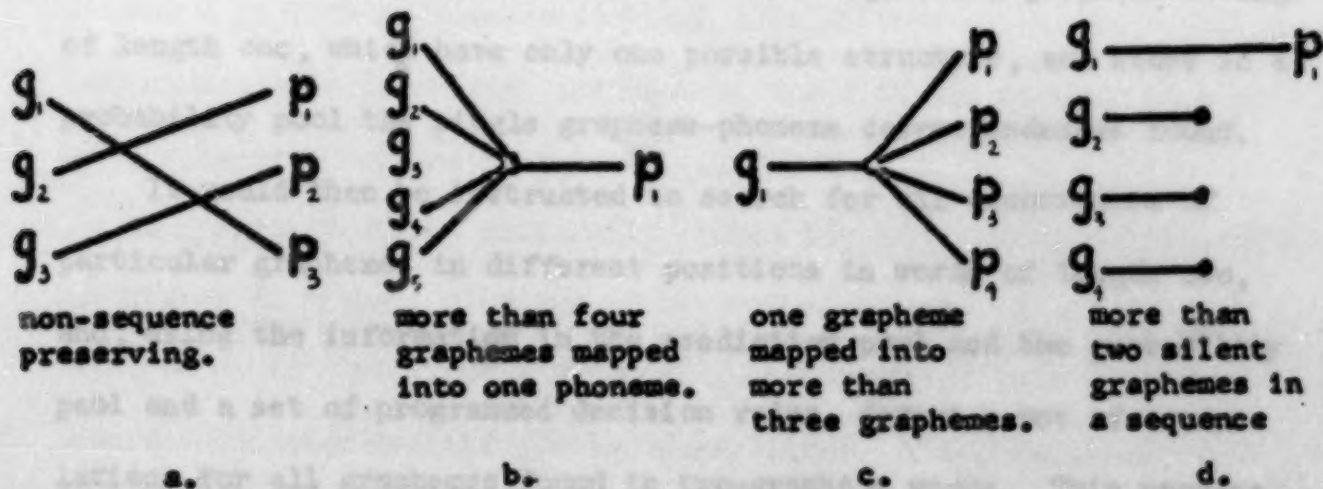


FIG. 3. ILLEGAL MAPPINGS

The phonemes, like the graphemes, are selected from a finite set and are combined into phoneme strings according to a set of construction rules for phonemes. The basic problem is to find a minimal description of the complete mapping.

If every possible mapping were possible, then we could go no further than describe, for every given grapheme string, the particular corresponding phoneme string. This is not true, however, in English. There are, as illustrated in Figure 3, page 6, a large number of mappings that do not occur in English. Those mappings shown in Figure 3 violate rules on what we will call the first level. This level contains all rules that can be stated without reference to particular phonemes or graphemes.

What we are looking for in considering these mapping rules is the minimum number of rules necessary for an accurate and complete correlation-deriving process. Given only the rules on level one, the problem can probably be solved. The procedure would be to program the computer to form for every phoneme-grapheme pair a prediction pool of all possible correlation structures, using only rules on level one. When this is completed, the computer would begin with grapheme strings of length one, which have only one possible structure, and store in a probability pool the single grapheme-phoneme correspondences found.

It would then be instructed to search for all occurrences of particular graphemes in different positions in words of length two, and, using the information in the prediction pool and the probability pool and a set of programmed decision rules, derive a set of correlations for all graphemes found in two-grapheme words. This process would then be repeated for three-grapheme strings, four-grapheme strings,

and so on, using the information derived in each case in the examination of the next graphemic word-length. Eventually this program should derive all spelling-to-sound correlations that exist in the words fed into it.

Unfortunately, we have no guarantee that such a system would always converge to the correct solution, and if it did we should suspect that the decision rules would need to be quite complex and the time required for convergence to be quite long.

1.7. The more practical approach is to use rules on level two, rules that apply to particular graphemes and phonemes, to reduce further the number of possible structures and then to design a computing routine that employs these rules to form correlations. This is the general approach that has been employed in this program. The rules on level two limit the particular phonemes that can correspond to particular graphemes. The computing procedure, which was derived by a trial-error process, forms alternate clusters of vowels and consonants in both the grapheme words and the phoneme words, and then finds the correspondences within each matching cluster-pair. Special routines determine whether W and Y are used as vowel-letters or consonant-letters in a particular word, whether a particular interconsonantal vowel-letter or an intervocalic consonant-letter is silent, and whether R and GH are members of vowel-letter clusters or consonant-letter clusters.

A special routine continually senses for any correspondences that are formed in violation of any of the rules on either level one or level two. If violations are found, they are stored in a special reserve and later printed out for examination. The program has also been designed to accept new rules of level two.

1.8. What the program does, therefore, is to start at the beginning of the grapheme word and form a cluster of all consecutive graphemes that are the same type (vowel or consonant). As soon as a grapheme is found that is not the same type as the grapheme before it, the program jumps to the phoneme word and repeats the process. When a pair of clusters has been formed, the program then jumps to a matching procedure that decides which graphemes within the grapheme cluster correspond to which phonemes within the phoneme cluster. This information is stored in a designated area and the entire process is repeated over and over until all graphemes and all phonemes have been matched. The entire set of correspondences are then stored and the process is repeated on the next dictionary word.

In actual operation this program will require two separate computer runs. In the first, correspondence for each grapheme-phoneme word pair will be formed and all correlations that occur for every grapheme will be tabulated and printed out. The correspondences and the correlation statistics will also be stored on magnetic tape. The second part of the program will be designed on the basis of the additional information that is desired after an examination of the results of part one. The major portion of the program for part two will probably be designed to tabulate different environments where particular correlations occur.

THE INTERRUPTABILITY OF WORDS IN THE SPEECH

OF NURSERY SCHOOL CHILDREN

A. L. Baldwin and Esther Baum
Cornell University

The primary focus of this study is the delineation of some of the underlying units in the language of children. In this study a unit of speech will be defined operationally in terms of its "interruptability." The assumption is that units are less vulnerable to being split in the middle of vocalization than non-units and the purpose is to discover such units.

SAMPLE

The investigation was conducted with children from the Junior and Senior classes in the Cornell Nursery School, who range in age from 3 to 5 years. At the time of testing, the 6 Seniors were between 4 and 5 years of age; the 6 Juniors were all more than 3 years, but less than 4 years of age, and represented one half of those children tested. The other 6 Juniors were unable to learn the task.

APPARATUS

The light in the experimental room emanated from a Santa Claus face lit up by a bulb within. When the switch was pressed the light went out, making a clicking noise, and throwing the room into darkness.

PROCEDURE

Ss were instructed to repeat a number of sentences after E. They were told that at some point during the repetition, the Santa would be turned out. At that time they were to stop in the middle of whatever they were saying, and "not make another sound." It was made clear that the light could go off at any time, and that they were to stop talking as soon as possible, even if this meant leaving words incomplete. Several

examples were given by E to words interrupted and split in the middle.

Ss were further told to "finish the sentence when the light comes on again." In cases where words were omitted, or where the tendency was to go back to the beginning of the sentence, Ss were repeatedly urged to "try to finish the sentence from where you stopped talking."

Trial sentences were used until each child reached the criterion level of interrupting three consecutive sentences successfully, that is, he stopped talking sometime before the end of the sentence.

Each of the 26 test sentences was composed of 11 syllables; the word or phrase to be interrupted was situated approximately in the center of the sentence. The following words were test words.

morning	2-syllable words, each interrupted in three different sentences, once just before the word began, once at the beginning and once in the middle.
even	
present	
follow	
2-when the	phrases, each interrupted in one sentence. The four labeled "1" were interrupted at the beginning of the first word and the four labeled "2" were interrupted between the two words.
1-they wanted	
2-play with	
1-came out	
2-see what	
1-show me	
1-some more	
2-to me	

The phrases were chosen because it was felt that they were common and familiar in the speech of young children and might function as units, that is, they would not lend themselves to being interrupted. Also, inspection of the record from the speech interval timer prior to testing showed that each of these phrases looked more like one long word, rather than a series of shorter ones.

<u>sunny morning</u>	phrases interrupted in one sentence each, and employing two of the 2-syllable test words. "sunny morning" was interrupted at the beginning of "sunny," "so even" was interrupted between the two words.
<u>so even</u>	

toys One syllable words, interrupted in one
new sentence each. Interrupted as early as
what possible in the word.

Poked interrupted once. This is a monosyllabic
word having three peaks of volume level.

The sessions were taped, and the test sentences later recorded through a speech interval timer onto a Brush inkwriting polygraph, an apparatus which yields a visible recording of momentary volume level. This helped to determine where the interruption signal came, where speech actually stopped, and how long the child took before stopping.

The following is a list of the sentences with the test words and points of interruption indicated.

1. Please make that side even to the other one.
2. Will you be able to follow the parade to town?
3. Do you know where the toys can be put tonight?
4. If it is a nice morning I'll go to the beach.
5. He was not present during the music lesson.
6. I will go upstairs when the game is over.
7. We know they wanted to play in the sandbox.
8. On a cold morning I dress very warmly.
9. Did he begin that new story before lunch?
10. Will they send even our brown doggy away?
11. That pussy cat will always follow the red ball.
12. Did you see the present that Johnny received?
13. I said I don't want to play with that doll.
14. If the sun came out we would have our picnic.
15. If you are present you can join in the games.
16. I know at what time the game really began.
17. They want us to follow Tommy inside.

18. When it is morning he'll wake up the turtle.
19. He doesn't even care what was decided.
20. You know I can't see what they are making here.
21. Will you please Show me where my jacket was put?
22. He will need Some more clay to finish the castle.
23. I hope you can give it to me right away.
24. Was it a sunny morning when you planted?
25. It's all so even I don't know what to fix.
26. We all know that he poked Mary in the back.

RESULTS

Only 6 of the Juniors were able to be trained to criterion level; the remaining 6 showed an inability to stop talking until the entire trial sentences had been finished, and these Ss were not tested further. All of the Seniors were able to perform the task.

Table 1 shows that for both Seniors and Juniors the 2-syllable test words tend to be completed regardless of where the "stop" signal comes. If the signal came before the test word some Ss stopped before the word; but once the word was begun there were only 27 out of the 110 cases where the word was not completed before stopping.

Table 2 represents a breakdown of the number of times each of the 2-syllable test words was split in the middle. The words are arranged in sequence from those that have more "natural breakage" in normal speech to those that have little or none (For example, in "present" a high "breakage" word, the volume level between the two syllables dips down almost to zero in normal speech,) There did not appear to be any relationship between the facility with which the words were split, whenever they were split, and their "natural breakage."

Tables 3 and 4 represent the resumption patterns, after the "resume" signal, for the two syllable words as a function of the point of cessation of speech. Inspection shows that in those cases where S was able to interrupt a word in the middle, he did not resume from the point of cessation of speech, but in most cases with the following word. In other words he skipped part of the word completely. Only in 2 of the 16 instances of word interruption by the Seniors, and none of the Juniors, did S resume at the point where he stopped, i.e., from the middle of the word. However, when speech cessation occurred at the end of a word, resumption was usually made with the following word in 22 of 25 cases.

Clear-cut exceptions to this general trend are found in two 3 year olds, Mark and Melissa, who, regardless of where they stopped, resumed by starting over again at the beginning of the sentence in 15 out of 18 instances.(see table 5).

Tables 6 and 7 show speech cessation when the 8 phrases were interrupted. In both Junior and Senior groups, the general tendency was to complete the entire phrase, regardless of where the stop signal came. In only 12 of 43 cases among the Juniors, and 5 of 39 cases among the Seniors was there a failure to complete the second words of the phrase. Of the 13 instances where the phrases were split in between words, 7 were on the phrase "see what."

On the four 2-syllable test words, only once out of 64 instances did a Senior fail to stop talking after the end of the interrupted word; however, these same Ss went on to complete the second word of a phrase 30 out of 39 times.

A similar comparison for Juniors shows that 11 of 56 times, they did not stop at the end of the 2-syllable words, but later on in the sentence.

However, on the phrases, 32 out of 42 instances showed cessation of speech after the second word in the phrase, or later.

Table 8 groups Seniors and Juniors together, and depicts the speech cessation in the two phrases "sunny morning" and "so even." These are phrases that employ in a different context two of the 2-syllable test words. In 9 out of 22 cases Ss completed the second word, in 8 cases they stopped at the end of the first word, in 2 cases after the following word, and in only 3 out of 22 cases did they interrupt in the middle of the word.

Table 9 represents speech cessation for Seniors on three 1-syllable words. Only once did anyone stop in the middle of the word; speech stoppage occurred twice before the word, and twice later on in the sentence. In the majority of cases (9/14) speech cessation occurred at the end of the test word. In table 10, a comparable analysis for the Juniors is found. There were no cases where the word was split in the middle by the younger children. In 3 out of 15 instances speech stopped before the word, in 8 instances upon the completion of the word, and in 4 cases later on in the sentence.

When the reaction times to the four 2-syllable test words interrupted at the beginning were compared with reaction times on the same words interrupted at some time during the words, the difference was found to be significant ($Z=2.01$, $p=0.0222$ using the Mann Whitney one-tailed test. $U=186.5$ Juniors and Seniors were grouped together $N_1=N_2=24$).

The difference in reaction times between phrases interrupted at the beginning and those interrupted in the middle was found to be significant at the .001 level ($X=0$, $N=10$ pairs, Juniors and Seniors grouped together).

The differences in reaction times between the phrase "sunny morning" interrupted at the beginning of "sunny" and the word "morning" interrupted at the beginning of "morning" are significant at the .05 level (Mann Whitney test, two-tailed, Juniors and Seniors treated together).

The average length of time taken for each S to say the 1-syllable test words was noticeably shorter than the mean time required to say the 2-syllable words. The difference is significant at better than the .001 level (sign test, one-tailed, $N=12$ pairs, $X=0$). However, reaction times on the three 1-syllable words interrupted at the beginning did not differ significantly from reaction times to the four 2-syllable test words interrupted at the beginning. It was found, though, that with the 1-syllable words, Ss tended more often to go on to the next word before stopping, thus increasing reaction times. There were 5/92 cases among the 2-syllable words where Ss continued to the next word before stopping, but 5/25 cases among the one-syllable words. This difference is significant at the .01 level ($\chi^2 = 8.3616$).

A comparison was made for each S between the mean reaction time to the 8 phrases, and his total reaction time. Only those phrases were used where interruptions came at the beginning, or simultaneous with the first word of the phrase. For the Junior group, the difference was not significant ($U=14.5$, $N_1=N_2=6$), $p=.30$. For the Senior group the difference is significant at $p=.032$ for a one-tailed test (Mann Whitney, $U=6$, $N_1=N_2=6$).

The purpose of these comparisons is to show that the fact the child is in the midst of saying a word prolongs the time required to stop. Thus a word not only resists interruption in the sense that it is not stopped in the middle, but also in the sense that it prevents the child from stopping until the word ends even though he has passed his typical reaction time.

Juniors had significantly higher reaction times than Seniors to the four 2-syllable test words interrupted at the beginning ($p=.032$, Mann Whitney test, one tailed, $U=7$, $N_1=N_2=6$). This might be accounted for by the fact that in 11 out of 56 cases, speech among the Juniors did not stop until one or two words after the signal came, whereas this only happened once in 64 cases among the Seniors.

The difference between the two groups in reaction times on the 8 phrases is not significant, $t=.260$, two tailed test ($U=14$, $N_1=N_2=6$, Mann Whitney test).

DISCUSSION

If one accepts the definition of a unit as a relatively uninterruptable sequence of sounds, the data indicate that for our older subjects, and some of the younger ones, words often function as units. Rarely was speech stopped in the middle of words. However, in those few cases where this did occur, speech tended to be resumed not from the point of cessation, but rather from the following word. That is, instead of the child's picking up from where he left off, in the middle of the word, he resumed talking with the next word in the sentence. One might speculate that as far as the child is concerned, the intention was complete, or the word was completed subvocally or through some other psychological mechanism, even though speech was audibly discontinued. If this is true, then it fits the hypothesis that the word is a unit. In any case it suggests that both the point of interruption and resumption tend to be unit boundaries.

One question concerns the "unitariness" of a unit. Are there reliable differences in the interruptability of different words? As far as the present study goes, there is some indication of such a difference. One syllable words were broken in the middle about 7 percent of the time, whereas two syllable words were broken in the middle about 20 percent.

the time. On the other hand, the 2-syllable word with the clearest breakage in normal speech (present) was split less often than any, in fact about like 1-syllable words.

If it is the semantic value of the word that makes it interruptible, nonsense syllables should reveal the fact. We have no data on this hypothesis. The whole question must remain open for further studies.

Nor are words the only units of children's language. Our data demonstrate that Ss reacted to the 8 phrases in a manner similar to the way they reacted to the single test words, showing a general reluctance to stop talking until entire phrases were completed. The phrase "see what" proved an exception to this rule, and was split between words more than half of the time. Accordingly, the conclusion can be drawn that the bond between these words is not very great, and that this phrase does not form a particularly strong unit.

How is it that some phrases come to be treated as units? Possibly certain words follow one another quite frequently, as witnessed in our test phrases, and these words also have little or no pauses between them. Consequently, children (and especially non-reading children) tend to react to these combinations much as they do to single words. This may or may not imply a failure to discriminate the two separate words in the phrase from the totality of the phrase. But it is clear that within the structure of sentences, seven out of the eight phrases employed in this study were treated as units.

As one moves on to the younger children in the sample, it becomes increasingly evident that units can be something other than either words or phrases. It will be remembered that only one half of the 12 Juniors tested were able to learn to stop talking when the light went out. It could be that for these children, the entire sentence forms a unitary

entity, making it extremely difficult for them to chip it up into smaller parts.

It is of interest here to note the behavior of one three year old who was making an heroic effort to be silent when the signal was given. Incapable of doing this, he would clamp the palm of his hand onto his mouth. One had the feeling that he understood quite well what was expected of him, but yet he could not prevent himself from continuing until the whole idea or sentence had been uttered.

Some additional insights might be yielded by the unique behavior of two Juniors, Melissa and Mark. These children were able to stop talking when signaled; but when they resumed talking, in almost all cases they started back at the beginning of the sentence (this despite repeated urging by E to "start talking again where you stopped"). One might view these Ss as representing a kind of transitional stage, in which it is possible to break down the sentence while speaking, but in which it is very difficult, if not almost impossible, to complete it without going all the way back to the first word.

The very relevant question of reaction times has not been dealt with thus far. Is the child stopping at the end of "real units," or does his behavior only show that it takes a certain time before he is able to stop vocalizing at all? That is, his reaction time happens to coincide with the length of time needed to complete a word, and this is why he stops at the ends of words.

Our data on reaction times, however, do not support the idea that the length of time taken to stop is the same regardless of what the child is saying. For instance, in those cases where interruptions came late in the words or phrases, reaction times were significantly shorter than when interruptions came earlier in the words or phrases. Thus, although

reaction times were seen to vary significantly, the unities of speech were almost invariably preserved.

A clear-cut illustration of this observation is to be found in the comparison made in reaction times between "morning" and "sunny morning." In one sentence, the word "morning" belonged to a larger unit, the entity "sunny morning"; in other sentences, it stood more or less alone, not so tightly bound into a phrase. Analysis showed that the reaction times to "sunny morning" interrupted at the beginning of "sunny" were significantly greater than reaction times to "morning" interrupted at the beginning.

Although reaction times to 1-syllable and 2-syllable words did not differ significantly, this can be explained by the fact that on 1-syllable words, Ss tended more often to continue on to the next words before stopping than they did on two syllable words. Thus the differences in reaction times were minimized.

A comparison of Juniors and Seniors on the four 2-syllable words showed that Juniors took significantly longer to stop. However, Juniors tended more often than Seniors not to stop after the word signaled, but rather after the following word.

There was no difference between the Junior and Senior groups on reaction times to the phrases.

	Just before Word	During or after Next word	Two words later
Stops Before Word	30		
Middle of Word	24	11	
Complete Word	20	91	29
During or after Next word	5	1	
Two words later			1

Table 1. Speech Cessation in Four 2-Syllable Words, Each Repeated Three Times In Different Sentences as Function of Point of Interruption

S E N I O R S			
	Place Interrupted		
	Just Before Word	Simultaneous With Start of Word	During Word
S Stops			
Before Word	8		
Middle of Word	7	9	1
Completes Word	15	13	10
During or After Next Word	1		
J U N I O R S			
	Place Interrupted		
	Just Before Word	Simultaneous With Start of Word	During Word
S Stops			
Before Word	2		
Middle of Word	7	2	1
Completes Word	15	8	10
After Next Word	4	1	3
Two Words Later			3
T O T A L			
	Place Interrupted		
	Just Before Word	Simultaneous With Start of Word	During Word
S Stops			
Before Word	10		
Middle of Word	14	11	2
Completes Word	30	21	20
During or After Next Word	5	1	3
Two Words Later			3

Table 2. Analysis of Number of Times Each of 2-Syllable Words Was Split in the Middle: Words Arranged in Sequence from Those Having High Breakage to Those Having None.

Words	Number of Times Split	
	Seniors	Juniors
PRESENT-breaks most clearly	1/15	1/15
EVEN-breaks less clearly	5/12	0/13
FOLLOW-no breakage	6/18	6/10
MORNING-no breakage	5/17	2/17
TOTAL	17/62 or 27%	9/55 or 16%

Table 3. Speech Resumption in Four 2-Syllable Words As a Function of Point of Cessation of Speech

S's Resumption	S E N I O R S		
	Place Where Speech Stopped		
	Before Word	Middle Word	End of Word
Beginning of interrupted word	6	2	1
From point where speech stopped	0	2	
From beginning of following word	2	12	32

Table 4. Speech Resumption in Four 2-Syllable Words As a Function of Point of Cessation of Speech.

S's Resumption	JUNIORS - Omitting Mark and Melissa			
	Place Where Speech Stopped			
	Before Word	Middle of Word	End of Word	End of Next Word
Beginning of interrupted word		1		
From point where speech stopped				
From Beginning of Following word		2	22	8
More than one word after speech stop	1		3	1

Table 5. Speech Resumption in Four 2-Syllable Words as a Function of Point of Cessation of Speech.

MARK & MELISSA - 2 Juniors					
S's Resumption	Place Where Speech Stopped				
	Before Word	Middle of Word	End of Word	End of Next Word	Two or More Words Later
Beginning of Sentence	2	3	6	2	1
Beginning of Interrupted Word		1			
Beginning of Following Word		2	1		
More Than One Word After Speech Stop					

Table 6. Speech Cessation in Eight Phrases As a Function of Point of Interruption

SENIORS					
S Stops	Place Interrupted				
	Just Before Word	Simultaneous With Start of Word	Middle of Word	Between Words	During Second Word
Before First Word	4				
Middle of First Word					
After First Word	2	3			
Middle of Second Word					
Completes Phrase	12	6	3	5	1
Next Word After Phrase		2	1		

Table 7. Speech Cessation in Eight Phrases As a Function of Point of Interruption.

JUNIORS					
S Stops	Place Interrupted				
	Just Before Word	Simultaneous With Start of Word	Middle of Word	Between Words	During Second Word
Before First Word	2				
Middle of First Word					
After First Word	5	2			
Middle of Second Word	1				
Completes Phrase	9	5	9	1	2
Next Word After Phrase	2	3	1		

Table 8. Speech Cessation in Two Phrases "Sunny Morning" and "So Even" As a Function of Point of Interruption

JUNIORS and SENIORS Grouped				
S Stops	Place Interrupted			
	Before or Simultaneous With First Word	Middle of First Word	Between Words	During Second Word
After First Word	5		3	
Middle of Second Word	3			
Completes Second Word	5		3	1
After Next Word or Later		1		1

Table 9. Speech Cessation in Three 1-Syllable Words as a Function of Point of Interruption.

S Stops	SENIORS		
	Place Interrupted		
	Just Before Word	Simultaneous with Start of Word	During Word
Before Word	2		
Middle of Word	1		
Completes Word	4	3	2
During or After Next Word	1	1	

Table 10. Speech Cessation in Three 1-Syllable Words as a Function of Point of Interruption

S Stops	JUNIORS		
	Place Interrupted		
	Just Before Word	Simultaneous with Start of Word	During Word
Before Word	3		
Middle of Word			
Completes Word	7		1
During or After Next Word	3		1

THE DEFINITION OF UNITS IN LANGUAGE

George J. Suci

The purpose of this portion of the project was to initiate research aimed at the empirical definition of units in language. It has been assumed by others (e.g., C. E. Osgood, 1954, and G. A. Miller, et al., 1960) that the production and the comprehension of language requires some kind of an organizing process which combines linguistic elements into a smooth output in production, or separates the smooth flow into smaller sequences in comprehension. Although "organization" in language is poorly defined and the concept does not lead readily to experimental operations, there is general agreement that it involves the grouping of the sequential elements of language into "chunks" or units for purposes of processing. In speech production, for example, it seems highly unlikely that the selection of what is to be uttered is made phoneme-by-phoneme -- rather the decision unit is at a higher level involving groups of phonemes as in words, and these units may be phrases or whole sentences. The problem of this research is to define what constitutes a unit under a given set of conditions.

In order to avoid circularity in the identification of a portion of language as a unit, we need a set of non-linguistic criteria for units. Three non-linguistic characteristics of units were hypothesized. If a portion of language behaves in accordance with all three characteristics, some confidence may be placed in the validity of each of the characteristics in defining units. The characteristics were: (1) that a unit sequence of speech should act as a whole and should resist divisibility or fractionation into its elements or into smaller bits; (2) that decisions as to what comes next in speech are made at unit boundaries and since this process of deciding takes time, pauses are produced between units; and (3) that the unit whole may be described by something other than a listing of its parts -- i.e.,

it is recodable. These characteristics led to three operational variables which were incorporated into a set of experimental procedures. The variables were: pause in speech, forgetting, and "mapping".

Pause is simply a time during which there is no sound in an utterance. Pause varies in terms of its position in the sequence, its length in time, and its quality (e.g., there are linguistic stops, breath pauses, filled pauses, etc.).

Forgetting is measured in a standard fashion comparing pre- and post-interference trials. Forgetting of learned sequences of languages should be in unit chunks. This is based on the characteristic of indivisibility or resistance to fractionation. If a linguistic sequence is a unit, it should be forgotten or recalled as a whole -- i.e., its elements should not occur in isolation but only together with the others.

"Mapping" refers to the process of recoding. One sequence of language is said to be equivalent to another sequence if each element of the first has a corresponding element in the second and the number of elements in the two are equal. This condition permits a one-to-one mapping of one sequence onto another. If the linguistic elements are identical, then the equivalence is obvious and the mapping is easily done. If the language is different -- that is, if recoding occurred, in the sense that the same thing was said in different ways -- the problem of equivalence is immediately magnified and mapping is difficult. Mapping may then be possible if a chunk in one sequence can reliably be identified as equivalent or not equivalent to a chunk in the other sequence. These chunks are hypothesized to be units. They are derived from the judgment of an observer that portions of sequences are the same.

Procedures were designed to see if the three types of variables were related in sequences of speech under controlled conditions. In general, the subject was asked to repeat something in a tape recorder and then to recall this same utterance after some kind of interference was invoked. Pauses

were observed in the initial utterance, and forgetting and recoding were observed in comparing the initial with the post-interference utterance. If pauses bounded the same sequences of speech which were identified by the other operations, the three variables were to be considered as valid measures of units in language.

Three kinds of linguistic materials were pre-tested with both children and adults: (1) a series of sentences of varying length and complexity; (2) a series of word lists which varied in the number of words at a time in each list which were in an acceptable grammatical sequence; (3) short narratives which varied from 30 to 80 words in length. In each instance the subject was to repeat verbatim a sentence or a list or a story. As interference, the same task with new material of the same type was repeated.

Sentences were for the most part repeatable in the first trial and were only very minimally affected by interference. They also produced very little pausing. This was true with children as well as adults. Sentences were discarded.

Lists which varied in "grammaticalness" were used because it was thought that they would provide a means of controlling the recoding process. It was thought that the grammatical portions were much more amenable to recoding than the ungrammatical sequences and therefore the subject's coding would be much more easily observed. This would make the mapping operation more reliable. This seemed to be true, but the lists were uttered in a way so different from natural speech that the validity of this procedure was questionable and the lists were discarded.

Short narratives seemed the most satisfactory. With this procedure the subject learns the story to a given number of correct trials. He then learns another story after which he recalls the first. Mapping and pause measures are applied to sequences in all trials, including the recall trials. The correct trial after learning and the recall trial are compared to measure

forgetting. In addition, another source for observing units is provided by this procedure. This is given by the learning process if it is assumed that learning proceeds by the acquisition of units and not of single elements. That is, it is hypothesized that the subject learns this type of material by coding a whole sequence of the input and storing this sequence as a chunk. These chunks are units which should show the characteristics described above. They should be bound by pause, they should be absent or present in toto, and they should be mappable as units from trial to trial. If this is true, every trial provides information regarding units.

The narrative-learning method will be used full scale with adults and with children. An attempt is being made to develop a narrative which is adequate for both children and adults so that the results from the two age groups are more comparable. Pause is being measured with an oscillograph with input from a tape recorder. A great deal of time is being spent in trying to define as objectively as possible what rules are to be applied in the mapping process.

Summary. The goal of this research is to define the units of language. Three characteristics of units have been described and operationalized. Units should be bound by pause, they should resist fractionation and they should be recodable. Measures of these characteristics are being related in data obtained from both children and adults in a learning-interference-recall procedure.

References

1. Miller, G. A., E. Galanter and K. H. Pribram. Plans and the Structure of Behavior. New York: Henry Holt, 1960.
2. Osgood, C. E. (Ed.) Psycholinguistics. Supplement to J. Abn. Soc. Psychol., Vol. 49, No. 4, Part 2, October 1954.